South Carolina
Alternate Assessment
(SC-Alt)

South Carolina's Alternate Assessment, SC-Alt
Spring 2009 Operational Administration

Technical Report
Unedited Draft

American Institutes for Research
South Carolina Department of Education

Table of Contents

Index of Exhibits

Introduction

This report details the design, development, and spring 2009 operational test results for the South Carolina Alternate Assessment (SC-Alt). The SC-Alt assessment consists of four content areas: English language arts (ELA), mathematics, science, and social studies. The assessments are administered across three grade-bands: 3–5, 6–8, and 10.

The first chapter of this Technical Report describes the background of the alternate assessments in South Carolina, the format of the previous assessments, and the need for a new alternate assessment.

The second chapter is comprehensive in its scope and includes information on the design of the alternate assessment and the development of tasks and items to measure academic growth among students with significant cognitive disabilities. The field-test designs are also summarized in the second chapter.

Chapter 2 further reviews how the design of the alternate assessment is unique in that it uses a Student Placement Questionnaire (SPQ) to maximize the efficiency of teacher and student testing time by guiding the teacher to administer tasks at a complexity level appropriate for the achievement level of each individual student. A thorough review of the SPQ is presented in Chapter 2.

Chapter 2 also reviews how the design of the assessment allows for the development of a vertical scale by linking grade-appropriate tasks across grade levels and complexity levels within grades. A vertical scale presents many benefits to the assessment system; these benefits will be discussed later in this report in the sections on scaling and score reporting.

Chapter 3 details the spring 2009 operational test administration in ELA, mathematics, science, and social studies; test administrator training; use of the SPQ; measures taken to ensure the accuracy of scoring; and the maintenance of test security.

Chapter 4 describes the standard-setting procedures conducted to establish performance standards. The chapter includes a description of the Item-Descriptor (ID) Matching procedure, the goals of the standard-setting workshop, the composition of the standard-setting panels, the workshop activities, and the panels' recommended performance standards.

Chapter 5 reviews several technical topics, including analysis and scaling and the reliability of test scores. This chapter includes a description of the procedures used to calculate internal consistency reliability estimates and classification accuracy estimates.

Chapter 6 describes the score reporting system for SC-Alt and references the Individual Student (Family) Report (included in Appendix H) from which the summary reports are derived. The chapter provides a brief description of the score reports, their intended uses, and the information they contain.

Chapter 7 provides an overview of statewide achievement on the SC-Alt, based on the spring 2009 operational test administration.

Chapter 8 reports on content validity and convergent and discriminant validity topics as well as the validity of the SPQ.

Chapter 1: Development of Alternate Assessment in South Carolina

Overview of the State Assessment System

The South Carolina Assessment System includes the South Carolina Palmetto Assessment of State Standards (PASS), the High School Assessment Program (HSAP), and the End-of-Course Examination Program (EOCEP). These state-level assessments are required by the Education Accountability Act of 1998 (EAA) as amended May, 2008 and are aligned with the state's academic standards for each subject and grade level.

- PASS measures the performance of all public school students in grades 3 through 8 in the content areas of English language arts (ELA), mathematics, science, and social studies.
- HSAP measures the performance of high school students in ELA and mathematics and is used both as one criterion for eligibility to receive a high school diploma and as the primary source for reporting the federally mandated data required by the No Child Left Behind Act (NCLB).
- EOCEP is administered in gateway courses at the high school level. The physical science EOCEP examination is counted for participation purposes for NCLB reporting.

The EAA establishes a performance-based accountability system that includes all students. This act supports South Carolina's commitment to public education and a conviction that high expectations for all students are a vital component of improving academic education.

The goals of the state assessment system are as follows:

- ***increasing academic performance of all children and, ultimately, raising high school graduation rates;***
- ***implementing rigorous academic achievement standards that are aligned with the South Carolina curriculum standards;***
- ***improving instruction based, in part, on the implementation of these higher standards; and***
- ***using the results of challenging assessments that measure student performance relative to these standards.***

Another goal is to inform various audiences—teachers, school administrators, district administrators, South Carolina State Department of Education (SCDE) staff, parents, and the public—of the status of academic performance and of the progress of public school students toward meeting South Carolina's academic achievement standards.

The South Carolina academic standards form the basis for alignment across the state education system for district and school curricula, classroom instruction, units of study, and learning experiences. The academic standards are the basis for all assessments in the state assessment system, including alternate assessment.

Purpose of the South Carolina Alternate Assessment

The purpose of the alternate assessment based on alternate achievement standards is to capture and evaluate the performance of students who have traditionally been excluded from

statewide testing programs and to improve instruction for these students by promoting appropriately high expectations and the inclusion of these students in state accountability for district report cards and for Adequate Yearly Progress (AYP) reporting at the school, district, and state levels.

Description of the South Carolina Alternate Assessment

The SC-Alt is administered to students who have been determined by the individualized education program (IEP) team to be unable to participate in the general state assessments even with appropriate accommodations. It is an alternate assessment on alternate achievement standards to the PASS for students in grades 3–8 and the HSAP and Physical Science EOCEP for high school students.

The test is administered to students who meet the participation criteria for alternate assessment and who are of the ages of typical students in grades 3–8 and 10. Students who are ages 8–13 (the typical ages for grades 3–8) are assessed in ELA, mathematics, science, and social studies. Students who are age 15 (the typical age of students in grade 10) are assessed in ELA, mathematics, and physical science.

The SC-Alt consists of a series of performance tasks that are scored by the test administrator (teacher) as they are administered. The performance tasks are scripted activities, and each task contains four to eight related items. The items have a scaffolded scoring script to reduce the complexity of the item when students do not respond successfully on the first attempt. All items are linked to the South Carolina academic content standards through the South Carolina Alternate Assessment Standards and Measurement Guidelines (ASMGs). The ASMGs are linked explicitly to the South Carolina academic standards for grades 3–8 and 10, although at less-complex or prerequisite levels. Currently there are ASMGS for science and social studies, but in the areas of ELA and mathematics, the ASMGs have been replaced by Extended Standards linked to the revised ELA (2008) and mathematics (2007) standards.

The SC-Alt has three forms: elementary, middle, and high school. Students are assigned to forms on the basis of their age on September 1 of the tested year. Students who are ages 8–10 are assigned to the elementary school form, students who are ages 11–13 are assigned to the middle school form, and students who are age 15 participate in the high school form.

The assessment is designed to minimize teacher and student testing burden by administering only those items that are well-suited to a student's achievement level. The test administrator completes a Student Placement Questionnaire (SPQ) to determine the most appropriate starting task for the student. Tasks are arranged in order of difficulty (from easiest to most difficult). Once the appropriate starting task is identified, test administrators continue to administer tasks until the student can no longer respond successfully.

The first operational administration of the SC-Alt was conducted during a seven-week testing window during spring 2007 in ELA, mathematics, and science. A census field test was conducted during the same assessment window for the social studies assessment. In spring 2009, an embedded field test was added. Documentation related to the 2009 operational administration is the focus of this Technical Report.

Background on Alternate Assessment Development in South Carolina

The 1997 amendments to the Individuals with Disabilities Education Act (IDEA '97) created the mandate to include all children, including children with significant disabilities, in state testing and accountability systems. The vision for the South Carolina alternate assessment system was initiated in early 1998 in response to the IDEA '97 regulations. This vision has driven the development and revision of alternate assessment in South Carolina.

A core team of staff from the SCDE Offices of Exceptional Children, Assessment, Research, and Curriculum and Standards met in March 1998 to develop a plan for designing an alternate assessment to meet the IDEA mandate and to be included in the state assessment system. The team's first steps were to convene a steering committee and seek technical assistance from the Mid-South Regional Resource Center (MSRRC) to explore strategies for designing an alternate assessment.

The Alternate Assessment Steering Committee was convened on May 12, 1998, to assist SCDE in determining how to include students with significant cognitive disabilities in statewide assessments. The committee comprised parents, special education and general education teachers, administrators, and representatives from other agencies. Dr. Ken Olsen of MSRRC provided the committee with technical assistance, including information on IDEA requirements, examples of options that some states were using or considering, and research available on alternate assessment. He facilitated a process that allowed the Steering Committee to reach shared foundational beliefs, address eligibility criteria and content and performance standards, and develop plans.

To ensure that all students, including students with significant disabilities, are included in the testing and accountability systems and have appropriate access to instruction in the South Carolina academic standards, the Steering Committee determined that the alternate assessment would be based on the following principles:

- All children can learn, be expected to meet, and be challenged to meet high standards.
- Special education is an extension and adaptation of the general education program and curriculum, rather than an alternate or separate system.
- The South Carolina State Board-approved standards are the foundation for all students, including students with unique needs and abilities.
- Measurement and reporting must be defensible in terms of feasibility, validity, reliability, and comparability.
- Results of the state standards-based program must be used to improve planning, instruction, and learning.
- An alternate assessment is appropriate for the few students for whom the state assessment, even with accommodations, is not appropriate.
- The alternate assessment is designed for a diverse group of students and should be flexible enough to address their individual needs.

The committee articulated these goals for the alternate assessment:

- to provide evidence that students have acquired the skills and knowledge necessary to become as independent as possible;
- to document the student's performance and the performance of the programs serving the student;
- to merge instructional "best practice," instruction in state standards, and assessment activities; and
- to provide information in the development of curriculum that is responsive to the student's needs.

The Steering Committee created the following participation guidelines to guide IEP team decisions regarding students who should participate in the alternate assessment:

- The student demonstrates a significant cognitive disability and adaptive skills, which result in performance that is substantially below grade-level achievement expectations even with the use of accommodations and modifications.
- The student accesses the state-approved curriculum standards at less complex levels and with extensively modified instruction.
- The student has current adaptive skills requiring extensive direct instruction and practice in multiple settings to accomplish the application and transfer of skills necessary for application in school, work, home, and community environments.
- The student is unable to apply or use academic skills across natural settings when instructed solely or primarily through classroom instruction.
- The student's inability to achieve the state grade-level achievement expectations is not the result of excessive or extended absences or social, cultural, or economic differences.

NOTE: The term *significant cognitive disabilities* was added by the South Carolina Alternate Assessment Advisory Committee to the criteria after the passage of the NCLB December 2003 regulations on alternate assessment.

The Steering Committee recommended that the state develop a portfolio collection of evidence of student progress toward the South Carolina academic standards similar in design to the Kentucky Portfolio Alternate Assessment. The committee also recommended that SCDE prepare a Request for Proposal (RFP) for a contractor to develop the alternate assessment. Advanced Systems in Measurement and Evaluation Inc. (ASME), which later became Measured Progress, was awarded the contract. This company, along with the Inclusive Large Scale Standards and Assessment (ILSSA) project at the University of Kentucky, began work with SCDE on the design of PACT-Alt.

A work group was convened to define the domain for instruction and assessment. To ensure that the South Carolina curriculum standards were the foundation for all students, including students with unique needs and abilities, the work group developed adaptations of the curriculum standards. The work group comprised special education teachers, regular education teachers, parents, administrators, higher education personnel, representatives from community agencies, and SCDE personnel. The work group process, which was facilitated by staff from

MSRRC, focused on the prerequisite skills found primarily in the curriculum standards in prekindergarten through grade 2.

The work group affirmed that special education services must operate as an extension of the general education program and curriculum rather than as an alternate or separate system. The standards in this initial document were identified as concepts that every student, including students with moderate to severe disabilities, should know or be able to perform. These selected standards, which focused on skills that were deemed essential and attainable for every student, were directed toward the following goals:

- enhancing the quality of students' communication skills;
- improving the quality of students' everyday living;
- improving students' ability to function in society and promote in them an acceptance of and respect for self and others;
- preparing students for transition into adult living; and
- moving students toward independence, which may range from a level of self-care with assistance to total self-sufficiency.

The extensions were based on the state academic content standards in prekindergarten through grade 2. For each selected standard, examples of essential real-world performance skills were developed. The articulation of these performance skills was designed to provide the rationale for teaching the standards and to serve as guides for teachers and parents regarding what the skill "looked like" when a student demonstrated it. The committee specified that these performance skills could be accomplished in home, school, and community environments through a variety of individualized communication systems and might incorporate a variety of supports, such as physical assistance, physical prompts, verbal prompts, and technology. The document *The Extensions and Adaptations of the South Carolina Curriculum Standards for Students Participating in Alternate Assessment* became the focus of the portfolio assessment process, HSAP-Alt performance tasks, and the professional development training. In 2002, this document was revised and renamed the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment*, but it was still aligned to curriculum standards for prekindergarten through grade 2. This work was based on the IDEA requirements and the thinking at the time about how students with significant cognitive disabilities should be included in the general education curriculum and assessment.

Beginning with the 2000–01 school year, students in grades 3–8 who met the participation criteria for alternate assessment were assessed with the portfolio assessment, PACT-Alt. In 2003, a high school assessment, HSAP, which was designed to meet AYP requirements, was added to the state assessment system, and an alternate to HSAP was developed to measure student proficiency in ELA and mathematics. A Stakeholder Committee with expertise in high school instruction of students with significant cognitive disabilities and academic standards was convened to guide the development of the high school alternate assessment, HSAP-Alt. The committee recommended designing an assessment based on performance on a series of tasks linked to the state curriculum standards. The HSAP-Alt consisted of a series of scripted performance tasks in ELA and mathematics with scaffolded

administration and scoring procedures aligned with the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment*.

One critical piece of the development and implementation process of PACT-Alt and HSAP-Alt was the provision of intensive professional development related to standards-based instruction, much of it based on the work of Harold Kleinert and Jacqui Farmer Kearns. A resource for professional development was their book *Alternate Assessment: Measuring Outcomes and Supports for Students with Disabilities.* Professional development was essential to the implementation of the portfolio assessment because the teacher was responsible for teaching a student the content related to the academic standards, assessing the student's progress, and providing evidence of the instruction and progress in the portfolio. Prior to the implementation of the alternate assessment and the IDEA requirement to include students with disabilities in the general education curriculum, many students with disabilities, especially those with significant disabilities, and their teachers had been excluded from standards-based instruction and professional development related to academic standards.

Transition from PACT-Alt and HSAP-Alt to SC-Alt

After seeking input on the vision of a new alternate assessment on alternate achievement standards from the Advisory Committee and teachers who were conducting alternate assessment, SCDE wrote an RFP for the redesign or design of the alternate assessment system. The design was to be consistent with South Carolina's commitment to the instruction and assessment of students with significant cognitive disabilities and NCLB requirements. The focus was to be on grade-level academic standards. The new system was to address concerns related to teacher burden and time involved in assessment while supporting improved instruction based on state academic achievement standards. Extensive training for test administrators was to be integrated into the design of the assessment.

In September 2004, a contract was awarded to the American Institutes for Research (AIR) to assist the state in revising the alternate assessment. AIR managed the administration and analyses of the PACT-Alt and HSAP-Alt assessments during the 2004–05 and 2005–06 school years while developing the new alternate assessment, the South Carolina Alternate Assessment (SC-Alt), with SCDE.

American Institutes for Research

The American Institutes for Research (AIR) has more than 50 years of experience as a nonprofit organization dedicated to assessment, behavioral science, and educational research. Subcontractors for the project include Measurement Incorporated, a leader in the field of hand-scoring customized assessments and in printing, packaging, distribution, and retrieval services, and INSITE, a company with a long history of working with SCDE. AIR developed the South Carolina HSAP and the EOCEP programs and has enjoyed a successful collaboration with SCDE for a number of years.

Chapter 2: Test Development

The South Carolina academic content standards are the basis for alignment across the state for district and school curricula, classroom instruction, units of study, and learning experiences. The curriculum standards are the basis for the Palmetto Assessment of State Standards (PASS), the High School Assessment Program (HSAP), the End-of Course Assessment Program (EOCEP) and the alternate assessment. An initial step in the design of the new assessment was developing Assessment Standards and Measurement Guidelines (ASMGs).

Development of the Assessment Standards and Measurement Guidelines

In April 2005, a committee comprising South Carolina special education teachers, content specialists, SCDE staff, and AIR staff designed the ASMG document to support the new assessment development. The process involved extending the state academic standards in ELA, mathematics, science, and social studies in grade bands 3–5, 6–8, and 10 to be accessible to students with significant cognitive disabilities. This document replaced the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment.*

The ASMGs are the foundation for the development of the assessment tasks for the SC-Alt. The ASMGs in each content area are distillations of the essence of South Carolina curriculum standards in each grade level.

Each content area committee reviewed the large array of standards and prioritized those in grade bands 3–5, 6–8, and 10 that they deemed most important to students "now" and "in the future." They then reduced the complexity of these standards, while retaining the essence of the grade-level content knowledge and skills, to make the academic standards appropriate and accessible for students with significant cognitive disabilities. The committee was careful to address both the depth and the breadth of the academic standards and used professional judgment based on experience with the population and the content to determine the standards to be assessed. The resulting document provides the link to the grade-level standards and indicators in the state academic standards.

The measurement guidelines give task writers and teachers the specificity necessary to translate the assessment standards into assessment tasks and items and classroom instruction. A list of individuals who were involved in this process is included in each ASMG content document.

NOTE: The ELA committee recommended that the standards in the Research Goal not be included in the assessment standards. The rationale for this recommendation was that this goal is not tested to any great extent in PACT because this content is primarily taught and assessed at the classroom level. Committee members, however, indicated that the Communication Goal included standards that they deemed very important to this population and they recommended including assessment standards for this strand.

The State Board of Education adopted mathematics and ELA academic standards in August 2007 and May 2008. State Board of Education requirements call for replacing the high school physical science end-of-course assessment for all students with a biology end-of-course assessment. The adoption of these new standards, which occurred outside the cyclical review timetable, and the replacement of the physical science end-of-course assessment with the biology

end-of-course assessment have a direct impact on the ongoing schedule for developing additional tasks for the task pool.

During the 2007 and 2008 school years, committees of special educators and general educators met to extend the revised ELA and mathematics academic standards and the biology end-of-course standards. These documents were designed to provide specificity for instruction as well as assessment, so the committees extended all standards and indicators including those for non-tested grades. These documents referred to as the Extended Standards replace the ASMGs in ELA and mathematics.

Stakeholder Input into the Development of the SC-Alt

To ensure the validity of the overall assessment process, a great deal of time and effort was spent obtaining input from various sources, including the State Alternate Assessment Advisory Committee, classroom teachers, parents, and other agency personnel.

South Carolina State Alternate Assessment Advisory Committee

The State Alternate Assessment Advisory Committee meets quarterly to provide oversight to the SC-Alt. The committee includes members of the original Alternate Assessment Steering Committee and the High School Stakeholder Committee. The committee also includes parents, special educators, representatives of higher education, content specialists, special education directors, and district test coordinators. Additional members include representatives from the Department of Disabilities and Special Needs, the University of South Carolina School of Medicine, the South Carolina Assistive Technology Project, the South Carolina Interagency Deaf-Blind Project, the Autism Society of South Carolina, and Pro-Parents of South Carolina.

The Advisory Committee provided input on its expectations for the revised alternate assessment during the first meeting with the contractor, AIR, on November 5, 2004. SCDE and AIR staff reported each step of the development process to the Advisory Committee at each meeting and sought its advice and recommendations.

Early Development Activities

At the recommendation of the Advisory Committee, AIR item writers visited classrooms in South Carolina during January and February 2005 to observe teaching strategies and materials that were in use. They also reviewed PACT-Alt portfolios for examples of evidence that teachers used to demonstrate progress toward proficiency on grade-level standards and examined the characteristics of the HSAP-Alt performance event in order to build on the existing system.

Teacher focus groups convened during January 2005 obtained feedback from teachers on the types of tasks they believed were appropriate, the protocol format they preferred, and the materials they recommended for inclusion in the assessment.

Qualified item writers employed by AIR were trained to write tasks and items specifically aligned with the alternate assessment standards and measurement guidelines. Item writing teams included AIR staff with expertise in the content areas; alternate assessment specialists; and consultants in the areas of instruction of students who are blind and visually impaired, students who are deaf and hard of hearing, and students with cognitive disabilities.

On February 14, 2006, prior to the development of science and social studies tasks, SCDE staff and the AIR alternate assessment specialist provided additional training to the writing teams. The training was based on *Designing from the Ground Floor*, materials developed by the National Alternate Assessment Center (2005).

Consideration of universal design was a focus throughout the development process. Items including passages and response options were developed to use objects, pictures, picture symbols, words, and numbers. Several tasks in all four content areas and at different levels of complexity were piloted with South Carolina teachers and students in March and May 2005. AIR staff then interviewed the pilot teachers to determine the item characteristics and parameters that teachers believed worked well or did not work.

Summary of the Development and Review of the Original SC-Alt Tasks

- The task and item development process began with the creation of task kernels. AIR was primarily responsible for the majority of task kernels, with input from SCDE and teachers in South Carolina. Task kernels are basic ideas for an assessment activity, stimulus materials, and purpose, which, based on their relation to the South Carolina ASMGs, were used to develop a task and its items.
- SCDE reviewed the task kernels and provided feedback to AIR on which kernels were acceptable, which were unacceptable, and which needed revision. These reviews included alignment with the ASMGs.
- AIR item writers developed the items and stimulus materials. These items were reviewed internally by the content experts for clarity, quality, and alignment with the ASMGs.
- Following the comprehensive AIR internal review, the tasks and items underwent technical review by AIR to ensure that the items were properly keyed and scaffolded, the instructions were appropriate, the stimulus materials were interpretable, and the items were generally consistent in design with other tasks and items under development.
- Items that passed internal review by the AIR development staff were reviewed by the senior content lead for each content area and the senior alternate assessment specialist. This review ensured that within the content area, tasks and items followed the design of the assessment and were consistent with respect to format, presentation, and general administration procedures.
- Before items were passed to SCDE, the project director reviewed all items to ensure that they were consistent with the foregoing factors across content areas and grade bands.
- Following the final internal AIR review, items were passed to SCDE for its review. During this process, SCDE staff, including content specialists, special educators, and assessment specialists, provided feedback to AIR on the design of the tasks and items, the alignment of items to the ASMGs, and the appropriateness of the items for use in South Carolina. Some items were revised by SCDE to improve alignment with the ASMGs.

- Approved items were placed into tasks for a small-scale tryout, conducted by AIR with the assistance of teachers in South Carolina and Northern Virginia and AIR staff. These tryouts provided invaluable information regarding the clarity of instructions, the utility of the stimulus materials, and the success of the items and tasks in producing expected responses. Items that showed obvious problems were revised or discarded.
- After changes were made to the prototypes as a result of the pilots and tryouts, a committee of South Carolina teachers was convened on July 12, 2005, to review the revised tasks and provide further input and recommendations.

Content, Bias, and Sensitivity Reviews

Once small-scale tryouts were concluded, AIR, SCDE, and educators in South Carolina reviewed the tasks and items for alignment with the ASMGs and for bias and sensitivity concerns. The content and bias and sensitivity review process was combined due to the direct impact of the task format, materials and language to the accessibility to the assessment for the population. Committees comprising teachers of students with significant cognitive disabilities, representatives of higher education, special education administrators, experts in the instruction of students with limited English proficiency, and content experts from across the state participated in these reviews to consider the following:

- alignment to the ASMGs and Extended Standards,
- bias for specific groups and types of disabilities,
- accessibility of the tasks to the entire population for whom the test was designed,
- specific characteristics of items that tend to exhibit bias or are not appropriate for or sensitive to the characteristics of student subgroups (e.g., exclusionary language, stereotypes),
- format and content of the tasks,
- accessibility of materials, and
- clarity of instructions and ease of administration.

The review committee meetings were conducted in November 2005, May 2006 and, for the Spring 2009 embedded field test, in November 2008. During the reviews, some items were recommended for revision or elimination by the committee members.

Development of Stand-Alone Field-Test Forms

- On the basis of the feedback from all the steps above, AIR conducted a final review and sign-off for all items and tasks. Following this review, the items and tasks were affirmed ready for field-testing.
- Prior to assembling tasks into field-test forms, the senior content lead for each content area and the project director reviewed the items and tasks one last time to determine whether the revisions were appropriate and maintained the alignment of the item to the targeted assessment standard and measurement guideline.
- Tasks and their items were then placed into field-test forms consistent with the specifications described earlier.

Item Data Review

- After field-testing, AIR and SCDE staff, including alternate assessment specialists, psychometricians, content specialists, and special educators, met to review the field-test statistics.
- They reviewed the statistics associated with each item and task to determine whether the items were functioning within expectations and whether the tasks were appropriately placed within the instrument. The statistical criteria applied to the field-test item data and to the operational item data are described in Chapter 5.
- The committee also considered teacher comments on specific items from the field test, data from field-test observations, and the results of the alignment studies to make decisions about the inclusion of items in the operational assessment.
- Items that did not meet these criteria were retained for possible future operational use (or were revised for recalibration).
- The Item Data Review meetings for the original independent field tests were conducted in August 2006 and June 2007, an Item Data Review for the Spring 2008 embedded social studies field test was held in connection with the planning meeting in September 2008. The Item Data Review meeting of the 2009 embedded field test items in ELA, mathematics, science and social Studies was conducted in July 2009.

Development of Operational Task/Item Pool

- AIR once again reviewed all data associated with the tasks and items to determine whether the items were functioning as expected and were useful for measuring the achievement of students in South Carolina.
- Items that survived all review and analysis criteria were placed into the operational task/item pool.

Design and Development of the 2006, 2007, and 2008 SC-Alt Field Tests

Following the task development process, the field-test forms were designed and produced. The primary purposes of the independent field-test administrations for English language arts and mathematics (spring 2006), science (fall 2006), and social studies (spring 2007) were to produce data to evaluate SC-Alt tasks and items and to guide the assembly of operational test forms to be used in 2007 and beyond. Student scores based on field-test data were not reported.

An embedded field test (spring 2008) tested the symbolate version of social studies task "George Washington", so that its performance could be compared against the text version used in spring 2007 field test.

The design, data collection, and analysis of the independent 2006 and 2007 field tests in ELA, mathematics, science and social studies and of the 2008 embedded social studies field test were discussed in the spring 2007 and spring 2008 operational technical reports.

Spring 2009 Embedded Field Test in ELA, Mathematics, Science and Social Studies

Beginning with the Spring 2009 administration, new field test tasks were administered along with the operational forms for the purpose of refreshing the operational item pool, to align the revised extended standards in ELA and mathematics, and to improve content coverage in science and social studies. Each subject on the operational form has twelve operational tasks, followed by two or three embedded field test tasks (see Exhibit 1).

Teachers were aware that they were administering field test tasks, but were encouraged in training not to treat them any differently from other tasks. Teachers were instructed to present the field test tasks to all students, no matter whether the students tested in the low-, medium-, or high-complexity ranges.

The field test items were calibrated on the same scale as the operational items. The IRT parameters, classical item statistics, fit and DIF statistics were subject to an item data review conducted with AIR and SCDE staffs, July 22, 2009.

Use of the Student Placement Questionnaires

The Student Placement Questionnaires (SPQs) are brief structured rating instruments that represent the range of communication levels and cognitive-academic functioning found in the population of alternate assessment examinees. AIR developed the SPQ for the South Carolina Alternate Assessment program.

The student placement process is intended to achieve several important goals:

- It matches student achievement levels with the difficulty of the tasks and items that are administered.
- It allows a maximum number of student item responses at an appropriate level of difficulty.
- It minimizes fatigue by targeting the assessment to the student.
- It supports the psychometric rigor of student scores. A student is administered a better targeted test than one that contains many items the student might find too difficult. Better test targeting contributes to better score reliability. Inasmuch as fatigue effects from the student's limited attention span are reduced, validity of the overall assessment is enhanced.

Teachers completed the SPQs in each content area to identify the most appropriate starting task for each student. For each subject, the SPQs prompted the teacher with between 12 and 15 "can do" questions (e.g., can this student recognize the sun, moon, Earth?). The questions were grouped by major content standards and sampled across low-, moderate-, and high-complexity levels. Each question rated the student's functioning on a 4-point scale, valued 0 to 3. Answering the 12 to 15 questions of each SPQ, summing the total score, and identifying the most appropriate starting task in a look-up table took test administrators approximately 6 or 7 minutes.

The look-up table identified ranges of SPQ scores that corresponded to one of three starting tasks. Teachers used the SPQs to assign students to starting points on the assessment. Cut points for the science SPQ were based on the rules derived for the mathematics SPQ but were altered for the number of items on the science SPQ. Details regarding the student participation, analysis, and conclusions drawn from use of the SPQ placement procedure appear below.

Administration: Placement and Stopping Rules

After teachers identified the most appropriate starting task for a student, they followed several rules as they administered the starting task and subsequent tasks. If starting at task 1, the teacher would administer at least five tasks; otherwise, at least seven tasks would be

administered. For detailed placement and stopping rules for the spring 2009 operational and field-test administrations, see Appendix B.

SPQ Summary

The preceding discussion reviewed some of the implementation procedures for the SPQ. Here we review two of the technical characteristics of the SPQ: the method used to select the SPQ recommended starting task and the usefulness of the SPQ as an indicator of student starting task.

The technical development of the SPQ and determination of the cut points to determine starting tasks are fully described in American Institutes for Research, 2008, *South Carolina Alternate Assessment (SC-Alt): Technical Report for English Language Arts and Mathematics Field Test Administration, Spring 2006*.

Usefulness of the SPQ for Determining the Starting Task. AIR has gathered information regarding the agreement between the SPQ recommended start points and the final observed start points by reviewing item data following each of the 2007, 2008, and 2009 administrations. The results of the study of 2009 data are reported in detail in Chapter 8.

Use of the SPQ pre-assessment score is only the first step in the procedure used by the test administrator in determining where the student should start the assessment. The instructions for using the SPQ include procedures requiring teachers to adjust the starting point below the SPQ recommended start point when the student is not successful on the first administered task. Alternately, after reviewing the assessment, some teachers may have judged that a student needed to start at a higher level than recommended by the SPQ. This result occurred almost exclusively when the SPQ recommended Task 1 as the starting point.

The results of the 2009 study indicate that the agreement between the SPQ recommended start point and the observed start point by content area were 92% for ELA, 91% for mathematics, 86% for science and 89% for social studies. Since the test administrator is required to make adjustments based on the student's success on the first task, and these adjustments are reflected in the agreement rates, the SPQ appears to be working very effectively for targeting the first task to begin the assessment process.

Teacher Scoring Accuracy

The design of the SC-Alt includes test administrator (teacher) scoring of student responses. The degree of accuracy with which the test administrator evaluates student performance determines whether the student receives the correct scores and the correct performance level.

A special video study was conducted during the 2009 administration to confirm that test administrators were following all scoring procedures accurately. For this study, scoring accuracy refers to the degree to which teachers follow scaffolding and scoring directions correctly and assign correct scores to student responses.

Scoring accuracy by the test administrators was evaluated by having trained raters at AIR review the videotapes of the test administrations and assume the role of the test administrator in scoring student responses. The AIR raters did not know the scores assigned by the test administrators at the time of their own assignment of scores. After the raters concluded their scoring of the student responses, the consistency between the test administrators and AIR raters was determined.

The scoring consistency analyses are presented in Appendix C. The results indicate that there was a high degree of consistency between the scoring of the test administrators and the AIR raters, suggesting that test administrators in South Carolina understood the scoring procedures and implemented them accurately when scoring student responses.

2009 Operational Test Booklets and Administration and Scoring Procedures

For each grade-band test form in each content area, tasks and items were selected that met the statistical criteria and that covered the breadth of the targeted ASMGs. The 2009 operational test forms in ELA, mathematics and science were unchanged from the 2007 operational administration, while the 2009 operational social studies form was the same as that of the 2008 operational administration. All operational forms had their tasks ordered by increasing difficulty as indicated by the empirical difficulty of the first item in each task, which was determined by Item Response Theory (IRT) analysis. The goal was to use technically sound assessment instruments to support valid inferences about what students know and can do relative to the ASMGs in each content area.

The SC-Alt operational administration in spring 2009 included three sets of test materials in English language arts, mathematics, and science: one for the 3–5 grade-band assessment, one for the 6–8 grade-band assessment, and one for the grade 10 assessment. The social studies assessment used two sets of materials for grade bands 3–5 and 6–8 (Grade 10 is not part of the social studies assessment). Teachers (test administrators) received a *Test Administration Manual* (*TAM*) and comprehensive training based on the manual and the test materials.

The 2009 test booklets contained 12 operational tasks in each content area. In each grade band, English language arts and mathematics had two or three embedded field test tasks added to the end of the form. On the science and social studies forms, sets of two or three field test tasks were appended for the elementary and middle school grade bands only. A more detailed overview of the assignment of field test tasks to grade band/content combinations is provided in Exhibit 1. Operational tasks were arranged in test forms in the order of the empirical difficulty of the first item in each task. Each test form (elementary, middle, and high school) included linking tasks to support psychometric linking of the grade-band score scales. Each task consisted of four to eight separate items. Teachers were instructed to administer a minimum of five or seven tasks to each student, depending on the SPQ designated starting point, and to continue administration of subsequent tasks until the student was no longer successful.

Teachers also received other materials with each test booklet:

- a manipulatives kit (with printed and physical manipulatives for all tasks),
- an Answer Folder for each participating student, and

- a Student Placement Questionnaire and directions for determining the starting task for each student.

Exhibit 1 summarizes the operational grade-band assessments and the numbers of tasks in each grade assessment for 2008.

Exhibit 1: Numbers of Operational and Field Tasks in Each Grade-Band Assessment, 2009

Grade Band	Total in Each Grade Band (Field Test Tasks in Parentheses)			
	ELA	Math	Science	Social Studies
3–5	12 (+ 3FT)	12 (+ 2FT)	12 (+ 3FT)	12 (+ 2FT)
6–8	12 (+ 3FT)	12 (+ 2FT)	12 (+ 2FT)	12 (+ 2FT)
10	12 (+ 2FT)	12 (+ 2FT)	12	

The approximate test length for each grade band assessment for the 2008 administration was 60 items (12 tasks × an average 5 items per task) and 120 score points (60 items × an average 2 points per item).

Linking Tasks in Each Grade-Band Assessment

All tasks in each SC-Alt grade-band assessment align with the original ASMGs in that grade band. For example, the first two items in Task 9, which is part of the grade band 3–5 ELA assessment (see Exhibit 2), align with Measurement Guideline 27, "Identify the problem and/or the solution in a story or drama," and Measurement Guideline 16, "Identify the impact of a given cause or effect on a given character." These Measurement Guidelines are linked to State Academic Standards and Indicators 3-R2.2 and 5-R1.11, respectively, from grade band 3–5.

Because adjacent grade-band score scales are linked psychometrically, some tasks in each grade-band assessment align with original ASMGs in both adjacent grade bands. For example, ELA Task 10, which provides data for psychometric linking of the grade bands 3–5 and 6–8 score scales, aligns with ASMGs at both grade bands. Similarly, item 1 in Task 10 aligns with ASMG 30, "Identify the purpose of a text"; that ASMG is linked to State Academic Indicator 3-R2.10 at grade band 3–5 and State Academic Indicator 6-R2.9 at grade band 6–8 (see Exhibit 2).

All items in linking tasks are designed to be appropriate for students in *both* adjacent grade bands. The alignment studies (discussed in Chapter 8) confirm that all tasks in each grade band, including linking tasks, align with ASMGs for each separate grade band and with the corresponding grade-band academic content standards. In addition, the corresponding grade-level State Academic Standards and Indicators to which the ASMGs are linked *do* differ across the adjacent grade bands. (See Appendix A.)

Exhibit 2: Two Tasks from the Grade Band 3–5 ELA Assessment That Illustrate the Alignment of Items of Grade-Band ASMGs and State Academic Standards

Item	SC-Alt ASMG		Corresponding State Academic Standards and Indicators from Grades 3–5		Corresponding State Academic Standards and Indicators from Grades 6–8
Task 9					
Item 1	Recognize conflict in stories: ▪ 27. Identify the problem and/or the solution in a story or drama.		3-R2.2: Demonstrate the ability to identify problem and solution in a work of fiction or drama.		--
Item 2	Determine cause and effect in texts read aloud or independently: ▪ 16. Identify the impact of a given cause or effect on a given character.	☐	5-R1.11: Demonstrate the ability to analyze cause and effect.		
Item 3	Analyze plots, characters, and settings in literature: ▪ 25. Identify and describe characters, settings, and events in a story.		3-R2.1: Demonstrate the ability to analyze characters, setting, and plot in a literary work.		
Item 4					
Item 5	Make predictions about text: ▪ 9. Use pictures and words to make predictions about texts read aloud or independently.		3-R1.7): Demonstrate the ability to make predictions about stories.		
Task 10					
Item 1	Identify the author's purpose: ▪ 30. Identify the purpose of a text.		3-R2.10: Continue identifying the author's purpose in a variety of texts.		6-R2.9: Demonstrate the ability to identify the author's purpose in texts in a variety of genres.
Item 2	Recall details: ▪ 7. Recall details in tests read aloud or independently.		3-R1.5: Demonstrate the ability to recall details in texts.		6-R1.5: Demonstrate the ability to identify the details that support the thesis of a particular text.
Item 3					
Item 4	Follow written directions:		3-R1.13: Demonstrate the ability to follow a logical sequence of written		6-R1.11: Demonstrate the ability to follow multistep

Item	SC-Alt ASMG		Corresponding State Academic Standards and Indicators from Grades 3–5		Corresponding State Academic Standards and Indicators from Grades 6–8
Item 5	▪ 17. Follow written one-step or multistep directions (presented in words/pictures/symbols/ objects).		directions to complete a task.		directions such as those for preparing applications and completing forms.
Item 6	Make predictions about text: ▪ 9. Use pictures and words to make predictions about texts read aloud or independently.		3-R1.7: Demonstrate the ability to make predictions about stories.		6-R1.6: Demonstrate the ability to make predictions about stories.

Chapter 3: Spring 2009 Operational Test Administration

This section describes the spring 2009 operational test administration:

- student participation for the spring 2009 administration;
- demographics of participating students;
- test administration window, materials, and timelines;
- test administrator requirements;
- test administrator training;
- pre-assessment using the Student Placement Questionnaire;
- fidelity of administration and accuracy of scoring; and
- test security provisions.

Student Participation for the Spring 2009 Administration

Students participating in the spring 2009 operational administration were those students whose IEP team had determined that they met the following SC-Alt participation criteria for alternate assessment and who were ages 8–13 or 15 on September 1, 2008. These are the ages of typical students who are in grades 3–8 and 10.

- The student demonstrates a significant cognitive disability and adaptive skills, which result in performance that is substantially below grade-level achievement; expectations even with the use of accommodations and modifications.
- The student accesses the state-approved curriculum standards at less-complex levels and with extensively modified instruction.
- The student has current adaptive skills requiring extensive direct instruction and practice in multiple settings to accomplish the application and transfer of skills necessary for application in school, work, home, and community environments.
- The student is unable to apply or use academic skills across natural settings when instructed solely or primarily through classroom instruction.
- The student's inability to achieve the state grade-level achievement expectations is not the result of excessive or extended absences or social, cultural, or economic differences.

Exhibit 3 indicates the age ranges of students who participated in the SC-Alt in spring 2009. Exhibit 4 indicates the alternate assessment eligibility categories that were placed in each eligible student's state precoding file. (Precoding files enabled SCDE and AIR to ensure that the appropriate SC-Alt assessment materials were delivered to teachers in time for the spring 2009 administration.)

Exhibit 3: Age Reference Sheet for 2008–09 Alternate Assessment, Spring 2009 Operational Administration

Age as of 9/1/08	Corresponding Birth Date Range		Test Required 2008–09	Precode AA Eligibility Code
	Beginning DOB	Ending DOB		
5	09/02/02	09/01/03	none	5
6	09/02/01	09/01/02	none	5
7	09/02/00	09/01/01	none	5
8	09/02/99	09/01/00	SC-Alt Elementary	2
9	09/02/98	09/01/99	SC-Alt Elementary	2
10	09/02/97	09/01/98	SC-Alt Elementary	2
11	09/02/96	09/01/97	SC-Alt Middle	3
12	09/02/95	09/01/96	SC-Alt Middle	3
13	09/02/94	09/01/95	SC-Alt Middle	3
14	09/02/93	09/01/94	none	5
15	09/02/92	09/01/93	SC-Alt HS	4
16	09/02/91	09/01/92	none	5
17	09/02/90	09/01/91	none	5
18	09/02/89	09/01/90	none	5
19	09/02/88	09/01/89	none	5
20	09/02/87	09/01/88	none	5
21	09/02/86	09/01/87	none	5

Exhibit 4: Precode Project Coding (Alternate Assessment Eligibility Field)

Code	SASI Dropdown List Description	Full Description
0	Criteria not met	The student does not meet criteria for alternate assessment.
2	**SC-Alt Elem School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt Elem School form** this current school year (8-10 years old on September 1, 2008.
3	**SC-Alt Middle School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt Middle School form** this current school year (11-13 years old on September 1, 2008.
4	**SC-Alt High School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt High School form** this current school year (15 years old on September 1, 2008).
5	**AltAssess NotAgeElig**	The student requires alternate assessment, **but does not meet the age eligibility requirements to be assessed with SC-Alt this current school year** (i.e., the student was younger than eight years, age 14, or older than 15 years on September 1, 2008).

Demographics of Participating Students

This section describes the demographics of participating students by test form (elementary, middle, or high school). Exhibit 5 presents the student demographics for participating students in each grade band.

For the purpose of this report, the inclusion of students was based on the same criteria applied in the reporting of student scores. A student was included if the following criteria were met: (1) a signed security affidavit was received for the student, (2) the student was not noted to be excluded from reporting for some other reason (e.g., inappropriate administration procedures), and (3) the number of coded responses met the attemptedness requirement for student scoring (i.e., five valid responses) in at least one content area. The population of students reported, therefore, includes 1,254 elementary school test forms, 1,093 middle school test forms, and 414 high school test forms.

According to the attemptedness requirements, a student's responses to a test form could be assigned to one of three completion status categories: completion ("student satisfied attemptedness rule"), invalid ("student did not satisfy attemptedness rule"), or not tested ("student did not answer any content area items"). For all content areas, the majority of students reported completed the administered test form; 99% or more of the eligible students completed ELA and mathematics, 68% to 69% completed science and social studies in the elementary and middle school grade bands[1], and 99% completed the high school science assessment. Of the remaining student records, typically 1% or less of reported test forms were categorized as not tested or not meeting the attemptedness criteria.

Given that the number of students to be assessed on the high school test form was approximately one-third the number of students assessed on either the elementary or the middle school forms, the proportion of demographic characteristics of the student population was relatively consistent across grade bands. In terms of ethnicity, African American students made up 51% of the assessed students consistently across grade bands; Hispanic students accounted for approximately 3% (2% to 4%); and White students accounted for 41% to 45% of the students across grade bands. Other ethnicities each accounted for less than 2% of the assessed population. Gender was also relatively consistent across grade bands, showing a slight proportional decrease of males from elementary and middle school to high school, with approximately a two-to-one ratio of male students (63% to 68%) to females (32% to 37%).

The classification of students in terms of English language proficiency was also consistent across grade bands. The majority of students (97% to 98%) were classified as "English Speaker II," meaning that they had never been coded as an ESL student. The remaining language proficiency classifications each accounted for less than 1% of students by grade band with the exception of "Pre-functional" (2% to 3%), indicating that the student scored "pre-functional" on the English language proficiency assessment and is receiving English as a second language (ESL) services. The percentage of pre-functional ESL students decreased across grade bands.

[1] Not all students were required to complete the science and social studies subject areas.

The grade reported for a student in the school's database is the grade reported for funding purposes (EFA grade) and is often determined by the location of the student's educational program instead of by the student's age or years in school. Therefore, approximately 8% of students administered the elementary form (for students ages 8–10, the typical ages of students in grades 3–5) had reported EFA grades lower than grade 3 or higher than grade 5, with most of these students classified in the contiguous grades of 2 and 6. Of students administered the middle school form (for students ages 11–13, the typical ages for grades 6–8), 24% of the students were reported at grades below grade 6 or above grade 8. The vast majority of these students were classified as grade 5 students (18% of all middle school form students), which indicates that these students were being served in educational programs housed in elementary schools. Of the students administered the high school form (for students age 15), 75% were reported as grade 9 or grade 10 (37% and 38%, respectively). Sixteen percent (17%) of the high school form students were reported as grade 8 students, indicating that these students were being served in educational programs housed in middle schools. The purpose of assigning SC-Alt grade band forms by age is to ensure that students are instructed and assessed on the appropriate grade band curricula regardless of where their educational programs are housed.

Fourteen different disability codes were reported for students assessed with the SC-Alt. The coding system allowed students to be coded with more than one disability code. Students with the primary disabilities of moderate mental disability, mild mental disability, severe mental disability, and autism made up 89% to 93% of the students assessed with the SC-Alt (89%, 91%, and 93% for the elementary, middle school, and high school forms respectively). Of these four primary disabilities classifications, the largest SC-Alt participation group was moderate mental disability (27-46%), followed by mild mental disability (21-28%), autism (14-22%), and severe mental disability (11-12%). Although a few students would have a primary disability of speech or language impairment, the vast majority of students received this code because they were receiving speech/language therapy as a supplementary service.

Exhibit 5: Summary of Demographic Information

	Elementary		Middle		High	
	N	%	N	%	N	%
STUDENT'S ETHNICITY						
African American	641	51.12	558	51.05	213	51.45
African American/American Indian	5	0.4	1	0.09	2	0.48
American Indian	3	0.24	5	0.46	.	0
Asian	14	1.12	10	0.91	3	0.72
Hawaiian/Pacific Islander	1	0.08	2	0.18	.	0
Hispanic	44	3.51	41	3.75	7	1.69
White	519	41.39	459	41.99	187	45.17
White/African American	16	1.28	11	1.01	1	0.24
White/American Indian	1	0.08	2	0.18	.	0
White/Asian	5	0.4	2	0.18	.	0
Other	5	0.4	2	0.18	1	0.24
Unknown	.	0	.	0	.	0
STUDENT'S GENDER						
Blank	.	0	.	0	.	0
Female	407	32.46	353	32.3	154	37.2
Male	847	67.54	740	67.7	260	62.8
ESL (LANGUAGE)						
Advanced	.	0	.	0	.	0
Advanced Waiver	.	0	.	0	.	0
Beginner	2	0.16	2	0.18	.	0
Beginner Waiver	.	0	.	0	.	0
English Speaker I	.	0	.	0	.	0
English Speaker II	1214	96.81	1059	96.89	406	98.07
Initially English Proficient	2	0.16	1	0.09	.	0
Intermediate	.	0	2	0.18	.	0
Intermediate Waiver	.	0	.	0	.	0
Pre-functional	35	2.79	23	2.1	8	1.93
Pre-functional Waiver	.	0	2	0.18	.	0
Title III First Year Exited	.	0	.	0	.	0
Title III Second+ Year Exited	.	0	1	0.09	.	0
Unknown	1	0.08	3	0.27	.	0
ELIGIBLE FOR FREE OR REDUCED-PRICE LUNCH						
Free	766	61.08	667	61.02	257	62.08
Reduced	114	9.09	82	7.5	40	9.66
No	374	29.82	342	31.29	116	28.02
EFA GRADE (REPORTED GRADE FOR FUNDING)						
1	4	0.32	.	0	.	0
2	64	5.1	.	0	.	0
3	426	33.97	5	0.46	.	0
4	447	35.65	24	2.2	1	0.24
5	288	22.97	196	17.93	11	2.66
6	16	1.28	344	31.47	5	1.21
7	7	0.56	334	30.56	2	0.48
8	1	0.08	154	14.09	71	17.15
9	1	0.08	21	1.92	153	36.96
10	.	0	9	0.82	156	37.68

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
11	.	0	2	0.18	13	3.14
12	.	0	2	0.18	2	0.48
COMPLETION STATUS: Student satisfied attemptedness rule						
ELA	1251	99.76	1092	99.91	414	100
Math	1248	99.52	1089	99.63	412	99.52
Science	865	68.98	751	68.71	411	99.28
Social Studies	850	67.78	740	67.7	.	0
COMPLETION STATUS: Student did not answer any content-area items						
ELA	3	0.24	1	0.09	.	0
Math	5	0.4	3	0.27	2	0.48
Science	387	30.86	342	31.29	3	0.72
Social Studies	399	31.82	347	31.75	414	100
Migrant Status	.	0	.	0	.	0
Home Schooled	.	0	2	0.18	.	0
Medical Homebound	10	0.8	19	1.74	6	1.45
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)						
Severe Mental Disability	156	12.44	120	10.98	49	11.84
Moderate Mental Disability	339	27.03	447	40.9	190	45.89
Mild Mental Disability	350	27.91	239	21.87	87	21.01
Autism	273	21.77	189	17.29	57	13.77
Deaf/Blindness	5	0.4	5	0.46	3	0.72
Emotional Disability	8	0.64	5	0.46	1	0.24
Hearing Impaired	21	1.67	20	1.83	13	3.14
Learning Disability	6	0.48	10	0.91	1	0.24
Multiple-Disability	104	8.29	90	8.23	37	8.94
Other Health Impaired	54	4.31	45	4.12	11	2.66
Orthopedically Impaired	62	4.94	46	4.21	14	3.38
Speech or Language Impaired	927	73.92	617	56.45	141	34.06
Traumatic Brain Injury	4	0.32	4	0.37	5	1.21
Visually Impaired	49	3.91	32	2.93	11	2.66
TOTAL	1254	100	1093	100	414	100

Test Administration Window, Materials, and Timelines

The following list presents important dates for the spring 2009 administration of the SC-Alt:

- SC-Alt test administration training for teachers new to the SC-Alt operational administration (did not administer in 2007 or 2008) – five regional SCDE workshops: January 12-16, 2009
- District-level SC-Alt test administration training for all test administrators: February 2 – 25, 2009
- Test materials arrived in district: February 25, 2009
- Assessment window: March 9–April 30, 2009
- Teachers returned materials to DTC-Alt: May 4, 2009
- Materials received by contractor: May 8, 2009

Teachers had approximately eight weeks to review the materials and complete the test administration. Teachers received both printed and physical manipulatives to use during test administration. They were also responsible for collecting a few common classroom items that were familiar to the student to use with several tasks.

Test Administrator Requirements

Test administrators are required to receive training on all phases of the administration of the SC-Alt and must be one of the following:

- a certified employee of the district;
- an employee of the district who is a critical needs teacher and has a letter of eligibility, an interim certificate, or a critical needs certificate;
- a substitute teacher who is certified and employed by the district on an as-needed basis;
- someone who was a certified teacher but has allowed the teaching certificate to expire owing to retirement, change of career, or some other reason and has been approved by the district test coordinator or the DTC-Alt as a qualified test administrator; or
- someone who is not certified but has been employed by the school district in an instructional capacity and has been approved by the DTC-Alt as a qualified test administrator.

If a test is administered in a location other than the school, the test administrator must meet the criteria specified above.

Test Administrator Training

Test administration training is required for all test administrators. The SC-Alt is individually administered with a standard script and scored by the test administrator as the assessment is conducted. Fidelity of administration and scoring is essential to the validity of the assessment results.

Teachers who administered the SC-Alt during the spring of 2009 who did not administer the SC-Alt in spring 2007 or spring 2008 were required to attend a South Carolina Department of Education (SCDE) training session. In addition, all teachers who administered the SC-Alt in spring 2009, including those who attended the SCDE workshops, were required to attend a district-level SC-Alt administration training session conducted by the district test coordinator for alternate assessment (DTC-Alt). At the completion of the training sessions, each test administrator was required to sign and submit to SCDE an acknowledgment of receiving training and readiness to conduct the assessment.

The training included the following elements:

- review of the eligibility criteria for students participating in the alternate assessment;
- overview of the ASMGs, emphasizing the link to the general education standards;
- explanation of how the assessment was developed, including the role of the review committees;
- review of test administrator requirements, test security, and test materials;
- training and practice in pre-assessment using the SPQ;
- description of the assessment format and procedures:
    - setup,
    - script,
    - scoring,
    - adaptive instructions;
- making SC-Alt tasks accessible;
- overview of assistive technology and the alternate assessment;
- administration and scoring instruction and practice using released test items provided on video clips of South Carolina teachers actually administering a task to students representing a variety of disabilities and ethnicities;
- scoring qualifying round; and
- review of procedures for receiving and shipping materials back to the DTC-Alt.

Pre-Assessment Using the Student Placement Questionnaire

As noted earlier in this Technical Report, the administration of the SC-Alt uses the SPQ as a pre-assessment instrument to determine the most appropriate starting point in the assessment. Recall that the SPQ requires the teacher to evaluate the student on 12 to 15 "can do" statements addressing the student's skills and knowledge in each content area on the basis of the teacher's prior instructional knowledge of the student. A total score computed from the teacher's SPQ responses indicates the initial starting task for the assessment. Once the assessment has begun, the test administrator is required to adjust the starting point for the student if the student is not successful on the first task. Rules have been established for adjusting the starting tasks and for determining when the assessment should be concluded. The starting and stopping rules used with the SPQs for the 2009 administration are presented in Appendix B.

Fidelity of Administration and Accuracy of Scoring

This section describes the steps taken to ensure the fidelity of administration and the accuracy of scoring.

During the assessment administration, a monitor must be present to observe all assessment sessions and verify the use of proper assessment procedures and the authenticity of student responses. Monitors must be trained and sign a Test Administrator Security Affidavit to verify that the appropriate procedures were used. The Test Administrator Security Affidavit contains a coded label to link the document to the student answer folder and includes the principal's verification of the use of appropriate assessment and scoring procedures. If this document is missing, the administration is considered an invalid administration.

AIR and SCDE conducted an audit of the spring 2009 administration and scoring by requiring school system staff to videotape a sample of SC-Alt administrations. A sample of students were identified for videotaping according to the following:

- All districts implementing the SC-Alt were required to videotape at least one student administration (all content areas).
- Each teacher included in the sample was required to videotape only one student.
- The total number of videotaped administrations per district was based on the number of teachers involved in the assessment for each district.

The test administrators of the sampled students were notified of their inclusion in the sample and were given instructions for completing the videotaping. Approximately 10% of all assessed students and 33% of all test administrators were sampled. The videotapes were returned to the contractor and scored by trained raters. More detailed information on this study is presented in Appendix C.

Test Security Provisions

This section describes the test security procedures associated with the SC-Alt. SCDE has the following test security measures in place:

- Each local school board must develop and adopt a district test security policy. The policy must provide for the security of the materials during testing and the storage of all secure tests and test materials before, during, and after testing. Before and after testing, all materials must be stored at a location(s) in the district under lock and key.
- Each District Superintendent must designate annually one individual in each district for each mandated assessment who will be the sole individual in the district authorized to procure test instruments that are used in testing programs administered by or through the State Board of Education. The designated individual for alternate assessment is the district test coordinator for Alternate Assessment (DTC-Alt). The DTC-Alt is responsible for receiving and distributing all SC-Alt materials and ensuring that all SC-Alt administration procedures and requirements are met.
- All school and district personnel who may have access to SC-Alt test materials or to the location in which the materials are securely stored must sign the Agreement to Maintain Test Security and Confidentiality before they are given access to the materials.

- Test administrators must be trained annually to administer the SC-Alt and must meet all test administrator requirements.
- An assessment monitor must observe all assessment sessions and verify the use of proper assessment procedures and the authenticity of student responses for each completed assessment.
- Test administrators must complete an SC-Alt Test Administrator Security Affidavit for each student they assess and submit the form with the student's assessment materials.

Chapter 4: Setting Performance Standards

This chapter describes the methods used to set standards on the SC-Alt assessments for the 2006–07 school year and beyond. It includes descriptions of achievement levels, the procedure used to set standards for each content area, the goals of the process, the composition of the panels, the workshop mechanics, and the standards set for each content area, including student impact information. Complete details of this procedure can be found in a separate report (*South Carolina Alternate Assessment Spring 2007 Standard Setting Technical Report*).

From June 25 to June 27, 2007, AIR convened a diverse panel of 105 educators and non-educators to recommend status performance standards based on the spring 2007 operational test administration data for ELA, mathematics, and science and field-test data for social studies.

Using the Item Descriptor (ID) Matching method (see Cizek & Bunch, 2007; Ferrara, Perie, & Johnson, in press), the panelists reviewed test items and the corresponding Descriptions of Achievement Levels (DALs) and then recommended performance standards for Level 2, Level 3, and Level 4 achievement levels. These standards were then translated into cut points on the student proficiency scale by AIR psychometricians. This section describes the process and outcomes of the standard-setting workshop.

Descriptions of Achievement Levels

DALs are key elements in standard-setting processes. DALs define the content area knowledge, skills, and processes that examinees at a performance level are expected to possess. The descriptions of Level 1, Level 2, Level 3, and Level 4 performance that SCDE developed make up the public statement about what and how much South Carolina educators want students to know and be able to do for each grade level and content area. In the ID Matching standard setting for the SC-Alt tests, panelists based their judgments on the DALs presented in Appendix D when they placed their cut scores.

The development of the DALs followed a multistep process involving AIR staff and SCDE staff working with committees of teachers, parents, and special education administrators. The process was begun by examining the DALS used with the other South Carolina assessment programs (PACT, HSAP, PACT-Alt, and HSAP-Alt) and the performance-level descriptors for alternate assessments used by other states. Following this preliminary study, a decision was made to draft four levels of descriptors, Levels 1–4, with the intent of having Level 3 represent "proficient performance" for NCLB reporting.

In the next step, staff examined item locations on the vertical scale for each grade band and looked for clustering of content strands and other response demands from the 2006 field tests in ELA, mathematics, and science. SCDE and AIR staff drafted DALS around these clusters as a starting point only. They then refined these drafts to ensure reference to all content strands and articulation within and across levels prior to submitting them for committee input.

A stakeholder committee met on March 30, 2007, and was charged with determining what proficiency "looked like" for students participating in each grade band of the SC-Alt. The committee reviewed the ASMGs, the draft labels for the DALs, and the draft DALs to develop refined DALs to recommend to SCDE. Members were instructed to consider descriptions that

allowed room for growth within grade bands and across grade bands and to recommend DALs that reflected an expectation that students will grow and achieve from year to year and demonstrate more knowledge. They were requested to design DALs to allow room for higher achievement.

After the meeting, SCDE and AIR staff reviewed these drafts to ensure consistency across grade bands and performance levels. The committee reviewed the DALs again on May 16, 2007, to refine them before the standard-setting workshop in June. Some additional refinement occurred during the standard-setting workshop. The official DALs were presented to the State Board of Education on September 12 and are posted on the SCDE website.

The ID Matching Standard-Setting Process

The ID Matching standard-setting process, described in the standard-setting plan submitted to SCDE and reviewed by the South Carolina Technical Advisory Committee, was used at a workshop in Columbia from June 25 to June 27, 2007, with a panel of 105 members. The panel was divided into four groups: an ELA group, a mathematics group, a science group, and a social studies group. Three tables in each workshop were assigned to anchor grade band 3–5. The other three tables were assigned to anchor grade band 10 (except in the social studies panel, which split the panel between grade bands 3–5 and 6–8). AIR staff provided training and led the participants through two rounds of ID Matching to first set the Level 3 standard and then the Level 2 and the Level 4 standards.

Before the participants made each of their recommendations using the ID Matching procedure, they were given a readiness form to ensure that they fully understood the task and were prepared to place the performance standard. Analysis of these evaluations showed unanimous agreement from the participants that they understood the task and were prepared to make performance standard recommendations.

Goals of the Standard Setting

The goals of the meeting, as stated to the panelists, were as follows:

- recommend performance standards on the ELA, mathematics, science, and social studies assessments that correspond to the DALs for Level 2, Level 3, and Level 4 performance levels;
- consider the agreement and impact data to guide judgments about item difficulty and placement of the performance standards; and
- recommend to SCDE the appropriate placement of cut points on the student proficiency scales for each grade-band assessment.

Panel Composition

The 105 panelists participated in recommending performance standards across four content areas: ELA, mathematics, science, and social studies. The overall composition of the panel followed the SCDE-provided specifications and was broadly designed to ensure that the panel was widely diverse and represented a cross-section of South Carolina's educators and non-educators. The composition of all panels is shown in Exhibit 6. The demographic breakdown of the standard-setting panelists appears in Exhibit 7.

Exhibit 6: Composition of the Standard-Setting Panels

	Panelist Role					
	Grades 3–5 Subpanel			Grade 10 Subpanel[1]		
Panel	**Table 1**	**Table 2**	**Table 3**	**Table 1**	**Table 2**	**Table 3**
ELA	Teacher - Special Education*	Administrator - Special Education*	Administrator - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*
	Curriculum Specialist-ELA	Curriculum Specialist-ELA	Administrator - Special Education	Curriculum Specialist – ELA	Administrator - Special Education	Curriculum Specialist – ELA
	Higher Education	Teacher - Special Education	Curriculum Specialist - ESOL	Teacher - Special Education	Curriculum Specialist – ELA	Curriculum Specialist - ELA
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education
	Teacher - Special Education		Teacher - Special Education		Teacher - Special Education	
Mathematics	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Administrator - Special Education*	Administrator - DTC*
	Teacher - Special Education	Administrator - Special Education	Curriculum Specialist - mathematics	Teacher - ESOL	Curriculum Specialist - mathematics	Curriculum Specialist - Autism
	Curriculum Specialist - mathematics	Curriculum Specialist - mathematics	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Curriculum Specialist - mathematics
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education
		Teacher - Special Education	Administrator –Principal	Curriculum Specialist - mathematics	Higher Education	Teacher - Special Education
Science	Teacher - Special Education*	Administrator - Special Education*	Administrator - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*
	Curriculum Specialist - science	Curriculum Specialist - science	Administrator - Special Education	Curriculum Specialist - science	Administrator - Special Education	Administrator - ESOL
	Teacher - Special Education	Teacher - Special Education	Curriculum Specialist - science	Teacher - Special Education	Curriculum Specialist - science	Curriculum Specialist - science
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education
	Teacher - Special Education		Teacher - Special Education			
Social Studies	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Teacher - Special Education*	Administrator - Special Education*
	Curriculum Specialist - SS	Administrator –Principal	Curriculum Specialist – Social Studies	Administrator - Special Education	Teacher - Special Education	Teacher - Special Education
	Teacher - Special Education	Curriculum Specialist – Social Studies	Parent	Teacher - Special Education	Curriculum Specialist – Social Studies	Teacher - Special Education
	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	Teacher - Special Education	

Notes (1) Grades 6–8 for Social Studies. (*) Denotes table leader.

Exhibit 7: Demographic Breakdown of Standard-Setting Panelists

		N
TOTAL		105
Gender	Female	94
	Male	11
School District	Richland 1	14
	Richland 2	10
	Horry County	5
	Lexington 2	5
	Dorchester County	4
	Florence 1	4
	Lexington 5	4
	Lexington 1	3
	Berkeley County	2
	Greenville County	4
	Charleston County	4
	Greenwood 50	2
	Kershaw County	2
	SC School for the Deaf and Blind	2
	Sumter 17	2
	Union County	2
	York 3	2
	Other School Districts (1 each)	32
Other	Department of Disabilities	1
	Pro Parents of SC	1
Race/Ethnicity	African American	19
	Hispanic	2
	White	80
	Unknown/Other	4
Position	Special Education Teacher	61
	Curriculum Specialist	23
	Special Education Administrator	12
	Administrator	3
	Higher Education	2
	ESL Teacher/Curriculum Specialist	3
	Parent*	4

*Three parents were also special education teachers and have been counted in both categories.

Standard-Setting Workshop Activities

Workshop participants recommended performance standards for the assessments during two rounds of deliberation for each DAL in each content area and in each grade band as follows.

- Set standards in anchor grade bands (3–5 and 10)
  - Participants complete Rounds 1 and 2 for each performance-level standard.
  - Table leaders *articulate* standards across grades and content areas (align them on the basis of content considerations).
- Set standards in intermediate grade band (6–8)
  - Participants complete Rounds 1 and 2 for each performance-level standard.
  - Table leaders *articulate* standards across grades and content areas (align them on the basis of content considerations and consistency with anchor grade standards).

The workshop agenda shown in Exhibit 8 shows the sequence of events for the three-day meeting.

Exhibit 8: Standard-Setting Workshop Agenda

Day	Approx. Times	Primary Activity	ELA	Mathematics	Science	Social Studies
1	8:00–11:00	Table leader training	24 table leaders (6 from each content area)			
	11:00–12:00 1:00–3:00	Panelist training and practice	105 panelists (23 to 29 from each content area)			
	3:00–5:00	Anchor grades, Level 3, Level 2, Level 4, round 1	14 panelists for grades 3–5; 13 panelists for grade 10	14 panelists for grades 3–5; 15 panelists for grade 10	14 panelists for grades 3–5; 12 panelists for grade 10	12 panelists for grades 3–5; 11 panelists for grades 6–8
2	8:00–9:00	Review Day 1, Finalize Round 1				
	9:00–11:00	Anchor grades, Level 3, Level 2, Level 4, round 2	14, 13	14, 15	14, 12	12, 11
	11:30–1:00	Anchor grades, moderation by table leaders, all content areas	24 table leaders			
	2:00–5:00	Grades 6–8, Level 3, Level 2, and Level 4, round 1	27	29	26	n/a
	2:00–5:00	Closing, final results, and evaluation for social studies	n/a	n/a	n/a	23
3	9:00–11:00	Grades 6–8, Level 3, Level 2, and Level 4,	27	29	26	n/a

Day	Approx. Times	Primary Activity	ELA	Mathematics	Science	Social Studies
		round 2				
	11:30–12:00; 1:00–2:00	Final moderation by table leaders, all content areas	18 remaining table leaders			
	1:00–3:00	Closing, final results, and evaluation	82 remaining panelists			

Throughout the week, the panelists had many opportunities to reflect on the pattern of performance standards they were recommending. Their general conclusion was that they were satisfied that the standards made sense from a content and experiential point of view. They felt that the patterns reflected the requirements of the content standards and the realities of student performance.

With few exceptions, panelists recommended standards that followed an orderly progression of increasing achievement across levels and grade bands. Specifically, with the exception of mathematics at the grade band 6–8 and grade 10, all recommended achievement-level standards increased in difficulty in subsequent grade bands. This fact is evident by examining the scale scores associated with each recommended cut score at each grade and achievement level. Exhibit 9 through Exhibit 12 show the scale score associated with the cut score recommended by each panel. These results were achieved through the process of setting cut scores at anchor grades, making sure they resulted in consistent expectations across grade bands, and providing articulated standards as a starting point for intermediate grade bands.

Cut Score Review and the Setting of Final Cut Scores

The results of the standard-setting workshops were presented to the Technical Advisory Committee (TAC) of the Office of Assessment, SCDE, on July 27, 2007. The TAC discussed the results of the standard-setting workshops, reviewed the articulation of the cut scores by grade level, and recommended strategies to the Office of Assessment staff for improving the articulation of the final scores while respecting and maintaining the basic cut score decisions made by the workshop panelists.

The Office of Assessment staff presented the following information to the TAC for review and discussion:

- demographic and disability characteristics of the spring 2007 test participants;
- spring 2007 standard-setting results from the standard-setting workshop conducted by AIR;
- score distributions and descriptive statistics; and
- collateral data, including results from PACT, PACT-Alt, HSAP-Alt, and HSAP.

In addition, Special Education Unit staff of the Office of Assessment stated two assumptions about the students assessed with the SC-Alt and their current levels of academic instruction:

- The vast majority of students with significant cognitive disabilities will improve in their academic achievement as a result of instruction, which will result in increased achievement performance across grades (i.e., from elementary through high school).

- Many teachers of students with significant cognitive disabilities are just beginning to implement academic standards-based instruction, and therefore the students assessed with the SC-Alt in spring 2007 have not received the level of instruction that is desired or expected in the future.

A consideration of the initial cut scores in light of these assumptions identified a need to improve the articulation of cut scores across grade levels, both in terms of the scale score growth expectations on the vertical achievement scale and the percentage of students identified in each achievement level.

The TAC recommended that SCDE staff consider adjustments to improve cut score articulation to be more consistent with expectations related to the design of the test and the achievement performance of the students. The TAC recommended that careful and thoughtful adjustments to the cut scores, guided by the standard error confidence intervals around each original cut, would be reasonable and acceptable policy adjustments.

A committee of Office of Assessment staff examined the scale score articulation and the percentage of students in performance levels by grade and recommended minor adjustments to the original cuts made by the workshop panelists. The adjustments made to each cut score and the resulting final cut scores are presented in Exhibit 9 through Exhibit 12.

The standard error of the panelist-recommended cut score in Exhibit 9 through Exhibit 16 was based on estimates of the standard error of the median suggested by Huynh (2003). However, two additional details about the standard errors of the median are important to note: First, the standard errors were based on the actual recommended cut scores, and any post hoc adjustment to the cut scores was treated as a constant adjustment. In other words, the adjusted cut score still had the same standard error. Second, the standard errors were initially calculated as standard errors of the page numbers in the ordered-item booklet. In other words, a standard error of the median equal to 2 meant that the error in the panelists' recommended cut score was about plus or minus two pages. The standard error was then transformed to the scaled score metric through linear interpolation. This was possible because each page number in the ordered-item booklet has a scale score associated with it.

Exhibit 9: Panel Recommended and Adjusted Final Cut Scores—ELA

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Grades 3–5					
Level 2	403	13.75	None	403	2.96
Level 3	466	9.54	None	466	1.59
Level 4	491	12.26	None	491	1.73
Grades 6–8					
Level 2	417	9.64	None	417	3.81
Level 3	473	7.99	0.5	477	1.09
Level 4	501	9.18	None	501	1.45
Grade 10					
Level 2	429	10.56	None	429	3.38
Level 3	478	9.11	1	487	0.66
Level 4	503	9.68	1	514	1.77

Exhibit 10: Panel Recommended and Adjusted Final Cut Scores—Mathematics

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Grades 3–5					
Level 2	423	10.22	-1	413	0.66
Level 3	476	9.59	None	476	0.21
Level 4	526	14.48	None	526	4.63
Grades 6–8					
Level 2	425	10.18	None	425	0.50
Level 3	476	9	1.5	489	0.16
Level 4	529	10.46	0.5	534	0.74
Grade 10					
Level 2	434	11.93	None	434	2.19

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Level 3	476	14.76	1.5	498	1.97
Level 4	528	13.19	1	541	3.82

Exhibit 11: Panel Recommended and Adjusted Final Cut Scores—Science

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Grades 3–5					
Level 2	430	10.83	None	430	1.51
Level 3	474	10.36	-0.5	469	3.25
Level 4	496	10.38	None	496	0.81
Grades 6–8					
Level 2	447	9.66	None	447	0.06
Level 3	484	9.61	0.5	489	0.50
Level 4	514	11.33	None	514	0.95
Grades 10					
Level 2	463	11.72	None	463	4.71
Level 3	492	14.44	1	506	8.13
Level 4	535	14.78	None	535	1.45

Exhibit 12: Panel Recommended and Adjusted Final Cut Scores—Social Studies

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores		
	Scale Score	Scale Score Standard Error of Measurement (± SEM)	Level of Adjustment (± SEM)	Final Cut Scale Score	Standard Error of Cut Scale Score
Grades 3–5					
Level 2	423	16.64	None	423	2.98
Level 3	485	14.39	0.5	492	11.93
Level 4	549	14	None	549	2.04
Grades 6–8					
Level 2	439	14.04	None	439	5.96
Level 3	490	12.58	1.5	503	1.28
Level 4	560	26.91	None	560	10.57

The final cut scores, the percentage of students performing at each performance level, and the cumulative percentage of students at or above each level are presented in Exhibit 13 through 16. The final cut scores were approved by the State Superintendent of Education and were presented to the South Carolina State Board of Education on September 12, 2007.

Exhibit 13: Percentage of Students at Each Performance Level—ELA

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	12.6 %	100.0 %
Level 2	403	25.4 %	87.4 %
Level 3	466	21.9 %	62.0 %
Level 4	491	40.1 %	40.1%
Grades 6–8			
Level 1	—	12.9 %	100.0 %
Level 2	417	23.3 %	87.2 %
Level 3	477	14.9 %	63.9 %
Level 4	501	49.0 %	49.0 %
Grade 10			
Level 1	—	13.4 %	100.0 %
Level 2	429	23.6 %	86.6 %
Level 3	487	12.5 %	63.1%

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Level 4	514	50.6 %	50.6 %

Exhibit 14: Percentage of Students at Each Performance Level—Mathematics

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	14.3 %	100.0 %
Level 2	413	30.8 %	85.7 %
Level 3	476	29.3 %	54.9 %
Level 4	526	25.7 %	25.7 %
Grades 6–8			
Level 1	—	15.9 %	100.0 %
Level 2	425	28.5 %	84.1%
Level 3	489	25.9 %	55.6 %
Level 4	534	29.8 %	29.8 %
Grade 10			
Level 1	—	16.1 %	100.0 %
Level 2	434	30.1 %	84.0 %
Level 3	498	28.9 %	53.9 %
Level 4	541	24.9 %	24.9 %

Exhibit 15: Percentage of Students at Each Performance Level—Science

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	19.8 %	100.0 %
Level 2	430	18.2 %	80.2 %
Level 3	469	17.5 %	62.0 %
Level 4	496	44.5 %	44.5 %
Grades 6–8			
Level 1	—	22.1 %	100.0 %
Level 2	447	18.5 %	77.9 %
Level 3	489	15.3 %	59.3 %
Level 4	514	44.0 %	44.0 %
Grade 10			

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Level 1	—	25.3 %	100.0 %
Level 2	463	25.0 %	74.7 %
Level 3	506	16.1 %	49.7 %
Level 4	535	33.6 %	33.6 %

Exhibit 16: Percentage of Students at Each Performance Level—Social Studies

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	19.3 %	100.0 %
Level 2	423	32.7 %	80.7 %
Level 3	492	30.1 %	48.1 %
Level 4	549	18.0 %	18.0 %
Grades 6–8			
Level 1	—	19.7 %	100.0 %
Level 2	439	27.3 %	80.3 %
Level 3	503	34.1 %	53.0 %
Level 4	560	19.0 %	19.0 %

Chapter 5: Technical Characteristics and Interpretation of Student Scores

This section describes the psychometric analyses conducted as part of the South Carolina Alternate Assessment (SC-Alt) 2007–2009 operational administrations. Only embedded field test items in the English Language Arts (ELA), mathematics, science and social studies assessment were calibrated and evaluated using 2009 operational data. However, in order to provide a complete description of the technical characteristic of the 2009 assessment in all content areas, this chapter also reports the data analysis results for the sections of the assessment which had previously been calibrated using the 2007 and 2008 operational data (see American Institutes for Research and South Carolina Department of Education, 2008, *South Carolina's Alternate Assessment, SC-Alt Spring 2007 Operational Administration* and American Institutes for Research; South Carolina Department of Education, 2009, *South Carolina's Alternate Assessment SC-Alt Spring 2008 Operational Administration*). The reported analyses are intended to ensure the quality of the items, the assessment materials and instruments, and the score reporting scales as measures of state academic standards.

As a reminder to the reader, there are three grade-band forms in each content area: elementary school (grades 3–5), middle school (grades 6–8), and high school (grade 10; excepting social studies). At each grade band, the assessments have three potential starting tasks that correspond to three levels of task complexity (high, moderate, and low). Students are assigned to a starting task on the basis of teacher judgments recorded in the Student Placement Questionnaire (SPQ) for each content area. Linking tasks connect the grade-band forms so that the vertical test scale could be created.

Analysis and Scaling of Items, Tasks, and Test Forms

The ELA, mathematics, science, and social studies assessments received comprehensive psychometric analyses, including initial item calibrations, after their earlier field-testing. Final calibrations were estimated for the ELA, mathematics, and science content areas on the basis of operational data gathered during the spring 2007 operational administration; final calibrations for social studies were computed from operational data of the spring 2008 administration. Calibrations based on operational data were considered superior to those based on field-test data. The vertical scales were also defined using the linking tasks as the vehicle that connected the elementary, middle, and high school forms.

AIR calibrated the operational items, estimated examinee proficiencies, and calculated scale scores and achievement levels for operational forms. This process entailed examining item statistics to ensure quality measurement across the range of the assessment, calibrating the items within each content area to a common scale, then applying a maximum-likelihood (ML) scoring algorithm to each student's responses to estimate his or her proficiency scores and assign the correct achievement level.

Assignment of Examinees to Starting Tasks and Item Calibration and Test Forms Linking

All eligible students participated in the spring 2009 test administrations. The case counts of near 1,200 students each in elementary, 1,100 in middle school and approximately 400 in high school, per content area, enabled effective calibration across task starting points and grade bands.

Students were assigned to one of three starting points on the basis of the sum of the teacher responses on the SPQ. The SPQ cut scores were shown to correlate with student achievement scores on the 2006 field-test administrations (for details, see American Institutes for Research, 2008, *South Carolina Alternate Assessment (SC-Alt): Technical report for English Language Arts and Mathematics Field Test Administration, Spring 2006*). The assignment of student starting tasks based on the SPQ cut scores was intended to expose students to items that were ideally suited to their current level of achievement while ensuring that (a) each student responded to an adequate number of items so that reliable and content-valid proficiency scores could be estimated and (b) an adequate number of students responded to each item for the joint calibration to be reliable.

Teachers were instructed to administer all tasks associated with the assigned starting point, with provisions for dropping to a lower starting point (task) if the student was unable to respond to the items in the task at the assigned starting point. Students who were assigned to high and moderate levels of the assessment but were unable to respond to items in the tasks at those levels may have been moved back to a less difficult starting point.

The linking design allowed a joint (concurrent) calibration of all items within a content area and the placement of the items on a common difficulty scale. The tasks actually used to link the grade-band forms (linking tasks) were selected, in part, on the basis of their moderate difficulty levels. Moderately difficult tasks contribute to more stable linking across levels than tasks that may be either too easy or too difficult for the examinees.

Linking across grade-band forms was accomplished by using common tasks across grade bands. Some of the tasks from the elementary form are on the middle school form; some of the tasks from the middle school form are on the high school form. (For details, see "Linking Tasks in Each Grade-Band Assessment" above.) In general, tasks are assigned in such a manner that the forms increase slightly in difficulty as examinees progress through the grade bands. This means that a task assigned to the moderate level of complexity in the elementary form may be assigned to the low-moderate level in the middle school form.

A similar linking design was employed for the social studies field-test forms.

See Appendix E for a summary of the linking design in each of the four content areas.

Analysis Plan

AIR's analyses presented in the remainder of this chapter were conducted in five steps:

1. data preparation and quality control,
2. classical item analysis,
3. review of items not meeting psychometric criteria for inclusion on operational forms,
4. joint calibration of items according to the Rasch model, and
5. final achievement estimation and scale score calculation for operational forms.

Data Preparation and Quality Control

Before analyzing the operational test data, AIR psychometricians performed a number of quality control procedures to ensure that scanning operations resulted in accurate data capture of the teacher-recorded student responses. Prior to the test administration, AIR verified all of the point values for each form's answer folder. For each form, two AIR staff members independently verified the possible responses and point values for each item.

After receiving the scanned test data, AIR analysts carefully examined the data file to verify its accuracy. Descriptive statistics were computed to ensure that student case counts on the pre-identification file generally corresponded to the actual counts based on test data at the state, school, and classroom levels. In addition, AIR verified that the total number of items in the data file matched the number of items on the answer folder and in the test booklet and then examined the frequency distributions of item responses to identify potential scoring problems, such as out-of-range values or unused response categories.

For purposes of item analysis and student scoring, respectively, non-response (NR) data were treated in two different ways:

For *item analysis and calibration purposes,* a student had to have at least three scored responses for the testing attempt to be considered valid. For a response to be considered a scored response, the test administrator had to have assigned a numeric score (0–4) to the student's response. If the administrator scored NR for all items in a task, the task was treated as not administered, and NR values were recoded as missing.

For *operational scoring* of student responses and estimation of student proficiency, however, the NR codes were treated as indications that the item was administered and that the student did not possess the content area knowledge and skill to respond. In this case, all NR values were recoded as zeroes and included in the student proficiency estimates. Following this recoding, tests were reexamined to determine the number of scored responses (0–4) in each content area. For operational scoring, a student had to have at least five scored responses of any kind for the assessment to be considered a valid attempt within a content area.

After the accuracy of the data file was verified, classical item analyses and Item Response Theory (IRT) analyses were performed. Several quality control procedures were taken to ensure the accuracy of these analyses.

As an initial step, the program control file was checked by two data analysts to ensure that form layout was correctly specified and that item response values were correct. As a second step, two analysts independently performed all analyses. Results of the parallel analyses were compared for mistakes by using commercially available file comparison software. Last, the analysis results were spot-checked by using other commercially available statistical software to ensure that the results were consistent across statistical software packages. *These comprehensive quality control steps are highly effective in detecting any issues that might influence the interpretation of the item analysis results.*

Classical Item Analysis

Classical item analysis for the SC-Alt operational and field-test forms was conducted using the *AM* statistical software (http://am.air.org). The item analysis yielded the proportion of students in each response category, the percentage of omitted responses for that item,[2] and the proportion of students who were unable to respond to the item because of access limitations (where relevant). Correlations between the item score and the test score were computed using adjusted polyserial correlations. For purposes of calculating item statistics, omitted items were treated as incorrect when there was at least one scored response within the same task (see above). Minimum and maximum point values, average item scores, and adjusted item-total polyserial correlations were calculated for all items.

Test form statistics, such as internal consistency reliability estimates and standard error of measurement statistics, were suppressed at this point because all students were not expected to take all items. Such statistics would be misleading before Rasch scoring was applied. Special marginal reliability analyses used to determine the reliability of the student score estimates are described in a later section of this chapter.

The proportion of students in each score-point category was calculated as defined by the item's scoring guidelines, as well as the proportion of students with blank responses within attempted tasks (i.e., those with at least one scored response). Item difficulty was computed as the mean score on the item across all students taking the form and with a scored response on that item. The average proportion of total points, calculated as the mean score divided by the total number of points possible on the item, serves as an additional measure of item difficulty.

Review of Items Not Meeting the Specified Psychometric Criteria

Classical item analysis provided information about the technical quality of the items; items failing to meet specified psychometric criteria were flagged for subsequent review. During field-testing of ELA and mathematics (spring 2006), science (fall 2006), and social studies (spring 2007), AIR reviewed all flagged items in concert with SCDE to determine whether they were of sufficient psychometric quality. For the 2007 operational forms in ELA, mathematics, and science, and for the 2008 operational form in social studies, AIR conducted a statistical review of the items to determine whether any operational items were performing in an unacceptable fashion. With the spring 2009 operational SC-Alt administration, AIR subjected all embedded field test items in ELA, mathematics, science and social studies to an item data review.

All item reviews were conducted using polytomous item flagging rules that AIR has put in place for other alternate assessments. Items were flagged for review for any of six reasons:

- Item-total score correlation was negative.
- The mean proportion correct did not increase for each successive score category.

[2] An item was considered omitted if no response was recorded for the item (or the test administrator marked NR on the student score sheet), but the student responded to subsequent items on the task.

- Item difficulty value indicated that the item was extremely difficult or extremely easy.
- The omit rate was high.
- Differential item functioning (DIF) was present.
- Item fit statistics were unacceptable.

Adjusted polyserial correlations were flagged if they were less than .20. Near zero or negative adjusted polyserial correlations may indicate a flawed scoring rubric, mis-ordering of response categories, reader difficulties in interpreting the rubric, or an item that does not measure the construct of interest.

Items were also flagged if the mean proportion correct (of items attempted) of students in a score-point category was lower than the mean proportion correct of students in the next lower score-point category. For example, an item was flagged if, on average, students who received 3 points on the item got a lower proportion of the total points possible on items with scored responses than students who received only 2 points on the item. This situation may indicate that the scoring guidelines are flawed. This flag was interpreted conservatively because students may take items with different average difficulty levels.

Items were flagged if the proportion of students in any score-point category was greater than .95. A very high proportion of students in any single score-point category may suggest that the other score points are not useful or, if the score point is the minimum or maximum score-point category, that the item may not be appropriate for students at that complexity level. Analysts must take into account the item-total score correlation as well as possible instructional factors when interpreting the statistic.

Items with omit rates greater than 5% were flagged. High rates of response omission may indicate confusion by test takers or administrators on how to respond to the item, excessive test speededness, or an item that was too difficult. It was expected that rates of omission may be somewhat higher for the alternate assessment population, so this was not considered a critical problem.

AIR conducted analyses of differential item functioning (DIF) on all test items to detect potential item bias. The purpose of these analyses was to identify items that may favor students in one group over students of similar achievement in another group. AIR conducted DIF analyses that compared African American/White and female/male student subgroups. The sample sizes were very small for each subgroup (see the demographic tables in the previous section); thus, DIF analyses have limited utility. However, DIF statistics were calculated for the purposes of item review.

For polytomous items, both the Mantel-Haenszel chi-square ($MH\,\chi^2$; Zwick & Thayer, 1996; Zwick, Donoghue, & Grima, 1993) and the Standardized Mean Difference (SMD; Dorans & Kulick 1986) were calculated. The classification rules are defined in

Exhibit 17. Items in the "C" DIF category, indicating evidence of differential item functioning on the items, were flagged for review.

Exhibit 17: Summary of DIF Classification Rules for Polytomous Items

DIF Category	Rule
C	The p-value of $MH\chi^2$ is less than .05 and |SMD|/|SD| is greater than 0.25.
B	The p-value of $MH\chi^2$ is less than .05 and |SMD|/|SD| is greater than 0.17 and less than 0.25.
A	The p-value of $MH\chi^2$ is not significant at the .05 level or |SMD|/|SD| is less than 0.17.

In addition, items were flagged on the basis of criteria set for INFIT and OUTFIT statistics produced by Winsteps. To evaluate item fit, we examined the item fit statistics provided by Winsteps. The mean square INFIT and mean square OUTFIT statistics reported by Winsteps are based on weighted and unweighted standardized residuals for each item response, respectively. These statistics indicate the discrepancy for each item between observed item responses and the item responses predicted under the Rasch model. Both fit statistics have an expected value of 1. Values substantially greater than 1 indicate unmodeled noise (model underfit), and values less than 1 indicate a lack of stochasticity (model overfit).

Because it is weighted by the variance of a person's response to an item, the INFIT statistic is sensitive to deviations from expected response patterns among high information items (i.e., items with locations near the theta estimate for the student) that could indicate structural problems with test items or test form construction (e.g., mis-keyed items, items not measuring the common underlying construct; conversely, excessive redundancy in item content resulting in overdetermined response patterns).

The OUTFIT statistic, an unweighted mean square, is sensitive to low information responses (e.g., easy items missed by high-ability students, difficult items correctly answered by low-ability students) and may therefore indicate the presence of outliers (Linacre & Wright, 1994). Items were flagged if the mean square INFIT or OUTFIT values were less than 0.7 or greater than 1.3. Misfitting items were evaluated in conjunction with the classical item analysis results to determine whether items should be included in the operational pool. We reviewed item fit as part of the scaling process and the item-data review process.

Items flagged on the basis of any of the aforementioned criteria were reviewed by AIR psychometricians and SCDE officials. First, a team of AIR psychometricians reviewed all flagged items to ensure that the data were accurate and properly analyzed, that response keys were correct, and that there were no obvious problems with the items. AIR recommended whether the item should be retained in the item pool or discarded, depending on the reason for the item flag and its effect on the quality of the assessment as a whole. SCDE had the final authority on whether the flagged items should be included in the operational scoring based on the item statistics and content appropriateness of the items.

In response to the item data review, one ELA field test item (ITS ID 1308) and one Science field test item (ITS ID 1261) were rejected because of poor item statistics; a second ELA

field test item (ITS ID 1431) was rejected because its rubrics treated *all* student responses as correct; and a third ELA item (ITS ID 1355) was modified and promoted to operational field test item status for the 2010 administration. Finally, a fourth ELA item (ITS ID 1407) was rejected by SCDE after the item data review. No mathematics or social studies items were found to violate the psychometric criteria so severely that they needed to be removed from scoring.

Item Response Theory Calibration and Linking Test Forms

This section describes AIR's procedures for item calibration using IRT techniques. Item parameters were estimated using the Partial Credit model (Masters, 1982) approach available using Winsteps software. A common item design was used to enable simultaneous calibration and linking across grade-band test forms in each content area. Items were jointly calibrated across grade bands in a single Winsteps run for each content area. This calibration approach put the item parameters of all grade-band test forms within a content area on the same scale.

For 2009, the results reported on the vertical scale appear in Exhibit 18 and Exhibit 19. It is interesting to note that the mean scores show a general upward trend across grades. This indicates that a vertical scale is a useful way to describe the results of this population of students. The exceptions are the mathematics and science performance in high school, which are essentially equal to the middle school results. Second, in almost every grade band, a few students were at the floor of the test (minimum scaled score equal to 260), but very few reached the ceiling (maximum scale score equal to 740).

Exhibit 18: Scale Score Statistics, by Grade Band, Overall

Subject	Statistic	Elementary	Middle	High
ELA	N	1252	1092	413
	Mean	490.1	508.8	510.3
	SD	72.75	81.37	81.21
	Min	260	260	260
	Max	718	707	729
Math	N	1248	1089	412
	Mean	489.15	507.1	498.2
	SD	71.94	82.5	71.85
	Min	260	260	260
	Max	707	705	718
Science	N	865	750	412
	Mean	488.57	506.9	501.3
	SD	72.13	80.16	77.73
	Min	260	260	260
	Max	740	729	740
Social Studies	N	852	738	
	Mean	497.2	516	
	SD	75.57	90.98	
	Min	260	260	
	Max	740	732	

Exhibit 19: Scale Score Statistics, by Grade Band, by Primary Disability

	Elementary					Middle					High				
	N	**Mean**	**SD**	**Min**	**Max**	**N**	**Mean**	**SD**	**Min**	**Max**	**N**	**Mean**	**SD**	**Min**	**Max**
ELA															
Overall	1252	490.10	72.75	260	718	1092	508.77	81.37	260	707	413	510.32	81.21	260	729
Severe Mental Disability	157	396.25	62.64	260	509	120	403.05	67.03	260	536	48	390.00	72.67	260	506
Moderate Mental Disability	338	482.49	45.80	260	718	446	501.27	53.37	260	707	190	513.10	50.59	321	729
Mild Mental Disability	350	541.06	55.40	373	704	239	583.76	65.74	441	707	87	588.69	47.79	465	729
Autism	272	486.17	61.10	260	718	189	496.42	60.88	260	649	57	481.02	77.87	260	654
Math															
Overall	1248	489.15	71.94	260	707	1089	507.10	82.50	260	705	412	498.21	71.85	260	718
Severe Mental Disability	154	396.56	62.81	260	519	120	401.37	63.34	260	518	49	391.35	64.45	260	488
Moderate Mental Disability	338	479.38	50.44	260	707	445	498.28	53.40	260	705	190	500.85	45.19	292	718
Mild Mental Disability	349	539.43	55.21	409	707	239	581.72	68.12	376	705	86	564.27	48.19	451	718
Autism	272	487.18	56.21	289	707	189	496.78	70.24	260	705	56	475.91	65.35	260	569
Science															
Overall	865	488.57	72.13	260	740	750	506.92	80.16	260	729	412	501.29	77.73	260	740
Severe Mental Disability	107	393.78	67.18	260	498	81	390.31	70.30	260	529	49	389.04	67.13	260	501
Moderate Mental Disability	221	483.24	55.83	264	740	315	502.88	52.82	260	667	189	507.63	49.69	260	680
Mild Mental Disability	249	534.66	47.38	391	733	170	575.71	55.77	427	729	86	571.97	47.41	472	740
Autism	187	485.09	55.74	267	740	123	495.80	64.13	260	729	56	469.91	74.06	260	638
Social Studies															
Overall	852	497.20	75.57	260	740	738	516.03	90.98	260	732					
Severe Mental Disability	103	397.07	68.28	260	499	88	402.83	73.12	260	575					
Moderate Mental Disability	237	491.81	58.59	260	740	308	514.67	62.95	260	667					
Mild Mental Disability	245	546.30	55.36	401	740	154	597.73	70.26	422	732					
Autism	182	493.16	60.88	338	740	117	489.45	70.36	260	732					

As an important feature, the South Carolina alternate assessment implements vertical scaling, permitting the measurement of student progress on the state content standards over time. Such a scale provides educators and parents with information that can be useful for monitoring student performance as students move through grades over time.

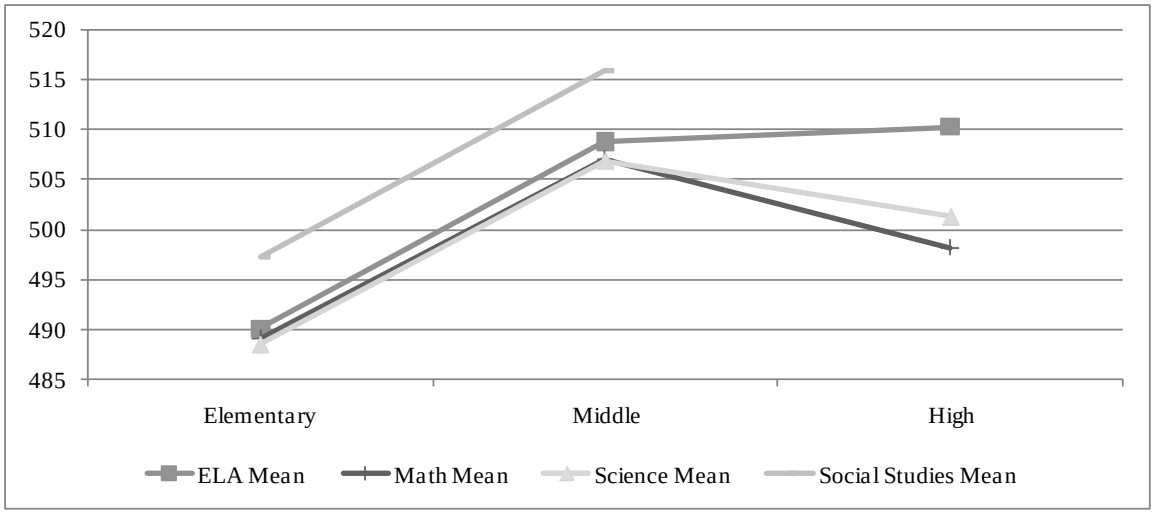
The development of this scale required the use of a common item linking design. In a common item design, *linking item tasks* appear on two adjacent grade-band forms of the assessment. These linking items allow for the grade-band scales to be connected, thus establishing the vertical scale.

There are at least two features of this linking design that warrant clarification. First, the linking items are the same (i.e., they are the same items) between two grade-band forms. They are connected to grade-specific standards in the higher grade as well as in the lower grade. As a result, students are never exposed to off-grade-level content since the common items serve a dual purpose in measuring content in both grade bands. Second, even though the same forms will be used the first two years of administration (pending planned new task and item development), it is not likely that exposure of the items to the students would trigger responses based on the recollection of any item's administration the previous year. As students grow academically, their starting task will likely change each year. New starting tasks mean that a portion of the items any student receives will be unique each year.

The South Carolina Department of Education is in the process of evaluating growth models under a federal grant (Modifications, Accommodations, Reports, and Standards). The use of the extant vertical scale for SC-Alt will facilitate deliberations for the possible implementation of a growth model for the alternate assessment.

A graph of the overall pattern of performance on the vertical scale is shown in Exhibit 20. Again, there is a general upward trend across all grade bands in each of the four subject areas. This graph shows that the vertical scale in the South Carolina Alternate Assessment was successful at capturing growth across grade bands.

Exhibit 20: Overall Pattern of Performance on the Vertical Scale



Using Item Responses to Estimate Student Proficiency

This section describes the estimation of student proficiency for the SC-Alt operational administration of English language arts, mathematics, and science assessments for elementary, middle, and high school; social studies assessments for elementary and middle school are also reported. The section first describes the estimation procedures used to determine student proficiency based on the items administered, then the transformation of proficiency estimates on the Rasch theta scale into scale scores, and finally the relation of achievement estimation to reliability estimation.

Student proficiency scores were estimated using a maximum-likelihood approach based on the scored items for each student.[3] This method calculates the theta score that maximizes the likelihood function of the given item responses for each student. Comparable scale-score estimates from these different item responses were achieved through the measurement-invariance property of IRT ability estimates, even when students were exposed to different ranges of items.

Under the Rasch-based IRT model, there is a one-to-one correspondence between the estimated theta score and the total raw score for a specific set of items. However, in the SC-Alt assessments, each student can take different sets of items. Using the pattern scoring method for calculating theta scores, we ensured that (a) two students who took the same items and achieved the same item scores were assigned the same theta score, and (b) students who took more difficult items were assigned higher theta scores than students with the same raw scores who took less difficult items. Thus, the scoring method took into account both the number of raw score points the student achieved and the difficulties of the items the student responded to. This scoring process was performed separately for each content area.

Once theta values had been estimated for each student, AIR converted the theta estimates to scale scores using a scale metric determined by SCDE in consultation with AIR. The SC-Alt assessments were scaled to have a mean of 500 and a standard deviation of 80 on the vertical scale for the grade band 6–8 assessment. The grade-band 3–5 and grade 10 assessment means and standard deviations were calculated in relation to grade-band 6–8 mean and standard deviation. This was done by performing a linear transformation of the Rasch theta scale for each content area, fixing the mean of the middle school test form scale at 500, and multiplying the student's theta deviation score by 80 as shown in the formula below:

$$y_{ijk}* = 500 + (\frac{\hat{\theta}_{ijk} - \hat{\mu}_k}{\hat{\sigma}_k}) * 80,$$

where

i indexes student;

j indexes grade band;

k indexes content area;

[3] The first step in this process was to rescore student responses consistent with the operational scoring method described under the "Data Preparation and Quality Check" section.

y^*_{ijk} is the scale score for student *i* in grade band *j* and content area *k*, given estimated ability, θ_{ijk};

$\hat{\mu}_k$ is the content-area-specific mean for the middle school test form; and

$\hat{\sigma}_k$ is the content-area-specific standard deviation for the middle school test form.

SCDE also decided to truncate the scale score ranges so that the lowest possible scale score was 260 and the highest possible scale score was 740. Student scale score estimates were truncated to the smallest whole integer (e.g., an estimated scale score of 440.60 would become 440). Additionally, scale scores were calculated and checked using a method similar to the process for total raw data.

Once scoring was completed, it was possible to estimate the internal consistency score reliability of the grade-band assessments by estimating the marginal measurement error across students. These estimates produced different standard errors for each student, depending on the items they were given and their level of performance on those items. Then, this value was used to determine the score reliability as the proportion of true score variance to observed score variance. We estimated this value within each content area (a) across the entire theta scale, (b) across grade-band forms, and (c) for each starting point within a grade band.

Test Score Reliability

This section provides the marginal reliability for each grade band, content area, and groups of students beginning at each starting task determined by the SPQ for the spring 2009 administration.

Classical test theory-based reliability indices, such as Cronbach's alpha, were not appropriate for the SC-Alt because the length of the test and the subset of items differed for each student. The reliability coefficient for the SC-Alt was, therefore, calculated as the *marginal reliability* (Sireci, Thissen, & Wainer, 1991), which is equivalent in interpretation to classical internal consistency estimates of reliability.

First we determined the marginal measurement error variance, $\overline{\sigma}^2_{e^*}$, across all examinees:

$$\overline{\sigma}^2_{e^*} = \int \sigma^2_{e^*} p(\theta) d\theta = \frac{\sum \sigma^2_{e^*}}{N},$$

where $\sigma^2_{e^*}$ is the square of the standard error of student ability estimate, $\hat{\theta}$. Thus, the marginal measurement error variance could be estimated as the average of squared standard error of $\hat{\theta}$.

Then we estimated the marginal reliability as

$$\overline{\rho} = \frac{\hat{\sigma}^2_\theta - \overline{\sigma}^2_{e^*}}{\hat{\sigma}^2_\theta},$$

where $\hat{\sigma}^2_\theta$ is the variance of observed θ estimates.

The marginal reliability estimate, $\bar{\rho}$, can be interpreted similarly to classical reliability indices such as Cronbach's alpha. Estimates of the marginal reliability for the test forms corresponding to the three SC-Alt grade-band assessments can be seen in Exhibit 21.

All marginal reliability estimates for each content area in each grade-band form met or exceeded 0.90, except for middle school ELA and science, both at 0.89, and middle school social studies at 0.88. The reliability estimates of all four content areas fall into the range of reliability coefficients found with large-scale assessments (Rudner and Schafer, 2001) and, subject to NCLB assessment system review, meet the reliability requirements for assessments used for the purposes for which the SC-Alt was designed.

In addition to the marginal reliability estimates, Exhibit 21 also displays the marginal standard errors of measurement for each subject and grade band, labeled $\bar{\sigma}_{e^*}$. These marginal standard errors of measurement range between 19 and 32 scale score units, placing the SEM at approximately one third of a standard deviation of the content area and grade band.

Exhibit 21: Marginal Reliability and Standard Error of Measurement by Grade Band and Subject

Subject	Grade Band	Elementary	Middle	High
English Language Arts	N	1252	1092	413
	Reliability	0.918	0.894	0.922
	$\bar{\sigma}_{e^*}$	20.859	26.521	22.633
Math	N	1248	1089	412
	Reliability	0.904	0.898	0.927
	$\bar{\sigma}_{e^*}$	22.307	26.359	19.431
Science	N	865	750	412
	Reliability	0.901	0.892	0.907
	$\bar{\sigma}_{e^*}$	22.713	26.315	23.679
Social Studies	N	852	738	x
	Reliability	0.901	0.875	x
	$\bar{\sigma}_{e^*}$	23.723	32.226	x

Appendix G shows the marginal reliability estimates broken out further by groups of students beginning at each starting task, as determined by the SPQ. The reliability coefficients in Appendix G are generally somewhat attenuated compared with those in Exhibit 21, due to the reduction in variance of scale scores grouped by starting task.

Classification Accuracy

This section describes the extent to which student achievement-level classifications were accurate across students. Classification accuracy was estimated for each cut score as the average probability of correct performance-level assignments across all examinees (assignments above or below the cut score), given each examinee's estimated proficiency score, θ_i :

$$\mathrm{CA_K} = \frac{\sum_{i=1}^{N_{k\geq K}} P(\theta_i > \theta_K^* \mid \theta_i, k_i \geq K) + \sum_{i=1}^{N_{k<K}} [1 - P(\theta_i > \theta_K^* \mid \theta_i, k_i < K)]}{N},$$

where

θ_i is the proficiency (i.e., theta) of student *i*;

k_i is the assigned performance level of student *i*;

θ_K^* is the cut score for the performance level *K* on the theta scale; and

N is the sum of the number of students at or above the cut score, $N_{k\geq K}$, and the number of students below the cut score, $N_{k<K}$, or simply the total number of students.

Thus, $P(\theta_i > \theta_K^* \mid \theta_i, k_i \geq K)$ is the probability that a student with θ_i assigned to performance level k_i is above the cut score, θ_K^*. The classification accuracy is the expected rate of correct classification probability, ranging from 0 to 1, where higher values indicate superior classification consistency. Exhibit 22 shows the classification accuracy by content areas, performance levels, and grade bands.

Exhibit 22: Classification Accuracy

Subject	Performance Level Cut Score	Grade Band			Overall
		3–5	6–8	10	
English Language Arts	Level 2	0.986	0.986	0.984	0.986
	Level 3	0.928	0.950	0.951	0.940
	Level 4	0.949	0.947	0.958	0.949
Mathematics	Level 2	0.979	0.974	0.978	0.977
	Level 3	0.931	0.947	0.929	0.937
	Level 4	0.942	0.951	0.947	0.946
Science	Level 2	0.959	0.968	0.956	0.962
	Level 3	0.943	0.945	0.923	0.940
	Level 4	0.927	0.941	0.926	0.932
Social Studies	Level 2	0.973	0.969	x	0.971
	Level 3	0.934	0.954	x	0.943
	Level 4	0.933	0.914	x	0.925

For example, according to the estimates in Exhibit 22, for the grade band 3–5 English language arts assessment 99% of students were correctly classified in Level 1 and 93% of students in Level 2 or above. All students in all grade bands had a probability greater than 0.923 of being classified accurately as proficient or higher (i.e., as level 3 or 4).

The calculation of probability of the correct performance level for students is described in the following section.

Calculation of the Probability of Being Classified Above a Cut Score Given the Student's Theta Score

For each student we can compute the likelihood of theta $L(\theta \mid \mathbf{z}, \mathbf{b})$. Suppose that the prior of the theta distribution is $f(\theta)$. Then, using Bayes' rule, we have

$$f(\theta \mid \mathbf{z}, \mathbf{b}) \propto f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}),$$

where $L(\theta \mid \mathbf{z}, \mathbf{b})$ is the likelihood of theta given the response $\mathbf{z}$ and item parameters $\mathbf{b}$; hence, the probability at and above cut is

$$P = \frac{\int_{\theta \geq \theta_{\text{cut}}} f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}) d\theta}{\int f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}) d\theta},$$

where $f(\theta)$ can take different distribution such as normal, or uniform, depending on our prior belief.

Calculation of $L(\theta \mid \mathbf{z}, \mathbf{b})$

For the Rasch model, we have

$$L(\theta \mid \mathbf{z}, \mathbf{b}) = \prod_{i \in MC} \left(\frac{Exp(z_i \theta - b_i z_i)}{1 + Exp(\theta - b_i)} \right) \prod_{i \in CR} \left(\frac{Exp(z_i \theta - \sum_{k=1}^{z_i} b_k)}{1 + \sum_{i=1}^{K_i} Exp(\sum_{k=1}^{i} (\theta - b_k))} \right)$$

$$\propto Exp(r\,\theta) \prod_{i \in MC} \left(\frac{1}{1 + Exp(\theta - b_i)} \right) \prod_{i \in CR} \left(\frac{1}{1 + \sum_{i=1}^{K_i} Exp(\sum_{k=1}^{i} (\theta - b_k))} \right)$$

where K_i is the maximum score for item i when this item is a CR item. It can be noted that the calculation above depends on total raw score r only when using the attempted items.

Chapter 6: Score Reports

This chapter describes the method used for reporting scores on the SC-Alt for the spring 2009 administration. An Individual Score Report (ISR) is included in Appendix H as an example of the highly detailed and diagnostic nature of the reports. This chapter gives a brief overview of how scores on the SC-Alt assessments are reported; a more detailed description is available in a separate *Score Reports User's Guide*.

The SC-Alt assessments have three types of score reports: the ISR, or family report; school reports; and district reports. Each report conveys specific information to its target audience. The reports are designed to be easily used by parents and educators. Of particular note, the reports include in-depth information about what students know and can do relative to the South Carolina academic content standards and to the performance levels.

The ISR provides specific performance feedback for each student across four content areas: mathematics, English language arts, science, and social studies (added in 2008). Within each content area, a graphic bar highlights the student's performance level along the proficiency scale. Each performance level is described in broad, easy-to-understand content terms. Further descriptions of what a student knows and can do are tailored and printed for each obtained performance level. For example, if a student is classified as Level 3 in mathematics, the following message is printed: "Students who score at Level 3 should be able to add and subtract simple numbers, count and compare objects in a group, compare objects by color, size, or shape, identify three-dimensional shapes, and read information in a graph." (Note: Scale scores were added to the ISR starting with the spring 2008 reporting cycle.)

Specific activities, based on each student's performance level for each content area, are presented for the family to do at home to help ensure positive academic growth in the content area.

The school report provides a summary of the performance of each student in the school. The alphabetical list of students contains basic demographic information and test form administered, in addition to achievement data. A scale score and achievement level are listed for each student for each content area. A school summary shows the number of students scoring at each performance level.

Three district-level reports are issued. The district roster summary report displays the roster of the district's tested students along with their demographic information, their scale scores and performance levels for each content area, and type of test form. The district summary by test form report presents a roster of schools in which students were tested, identifying the test form and giving the number of students tested in each content area and the percentages achieved in each performance level by content area. The total number of students tested with each form and their performance-level distributions by content area are listed at the bottom of the report. The district demographic summary report shows the number of students tested and the distributions of performance levels in all content areas, disaggregated by gender, ethnicity, lunch program, migrant status, and ESL status.

The separate *Score Reports User's Guide* has more specific information on how to interpret student scores and score reports and how to relate academic growth as measured by the SC-Alt to classroom curricula and activities. The guide has been widely distributed throughout South Carolina.

Chapter 7: Student Performance Data from the Spring 2009 Administration

Performance data from the spring 2009 administration are presented in this chapter. This was the third operational administration of the SC-Alt ELA, Mathematics and Science assessments and the second operational administration of the SC-Alt Social Studies assessment.

A total of 2,765 students from 84 school districts and 516 schools were tested with the SC-Alt in fall 2009. There are 2 student records that had neither school nor district information coded. The total number of tested students with one or more valid content area scores was 1,254 for the elementary form, 1,093 for the middle school form, and 414 for the high school form. Four of those students tested on two forms each and another four students did not have valid responses in any of the content areas.

Approximately one-half of the participating school districts (41; 49%) tested 15 or fewer students; 28 districts (33%) tested 16 to 50 students; and 15 districts (18%) tested more than 50 students each. Only six districts tested more than 100 students; the greatest number of students tested in one district was 237.

Of the 516 schools testing SC-Alt students, 346 (67%) tested five or fewer students; 113 (22%) tested six to 10 students; 46 (9%) tested 11 to 20 students; and 11 schools (2%) tested 21 or more. Only two schools tested more than 50 students each (61 and 65 students).

The elementary school form was developed to be administered to students who are 8, 9, or 10 years old at the beginning of the school year, which are the ages typical of students enrolled in grades 3, 4, and 5. The middle school form was developed for students of ages 11, 12, and 13 (typical of students enrolled in grades 6, 7, and 8), and the high school form was developed for students age 15 (typical age of students in grade 10).

Students tested with the elementary and middle school forms with reported ages outside the specified age ranges were either erroneously assigned to the forms by the test administrator or, in some cases, are the result of birth dates coding errors on the data files. The numbers of students reported outside the expected ages for the elementary and middle school forms is less than 1% for each content area. Students reported as having been tested on the high school form with ages below 15 appear due to form assignment or birth date errors. Students older than 15 (e.g., 16) are assessed with the high school form if they have not been assessed at the high school level previously.

The performance of students by grade-band form, age, and demographic group for the ELA, mathematics, science, and social studies content areas is presented in Exhibit 23 to Exhibit 26.

Exhibit 23: Performance by Grade-Band Form and Student Age—ELA

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	139	490	62	33	67	.	.	.	.	.	.	.	.	.	.
9	423	488	71	33	67	.	.	.	.	.	.	.	.	.	.
10	418	501	77	28	72	2	427	138	50	50	.	.	.	.	.
11	4	461	84	25	75	406	506	81	33	67	.	.	.	.	.
12	.	.	.	.	.	349	506	80	34	66	.	.	.	.	.
13	.	.	.	.	.	333	515	83	28	72	.	.	.	.	.
14	.	.	.	.	.	1	574	.	0	100	2	497	47	50	50
15	1	325	.	100	0	.	.	.	.	.	395	510	81	34	66
16	.	.	.	.	.	.	.	.	.	.	16	515	86	25	75
Blank	267	477	71	37	63	1	460	.	100	0	.	.	.	.	.
STUDENT'S ETHNICITY															
African American	641	495	76	30	70	557	514	83	31	69	212	517	74	31	69
African American/American Indian	5	483	34	40	60	1	384	.	100	0	2	595	108	0	100
American Indian	3	482	27	0	100	5	475	141	40	60	.	.	.	.	.
Asian	14	479	35	21	79	10	509	70	30	70	3	408	142	67	33
Hawaiian/Pacific Islander	1	353	.	100	0	2	512	9	0	100	.	.	.	.	.
Hispanic	44	466	63	45	55	41	499	74	27	73	7	502	85	29	71
White	517	487	69	33	67	459	502	78	34	66	187	504	87	37	63
White/African American	16	490	99	31	69	11	542	112	27	73	1	492	.	0	100
White/American Indian	1	646	.	0	100	2	531	122	50	50	.	.	.	.	.
White/Asian	5	464	43	40	60	2	507	35	0	100	.	.	.	.	.
Other	5	438	71	60	40	2	606	143	0	100	1	493	.	0	100
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
STUDENT'S GENDER															
Female	407	480	76	35	65	352	509	88	30	70	154	513	83	33	67
Male	845	495	71	31	69	740	508	78	33	67	259	509	80	34	66
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															
Pre-functional	35	466	67	43	57	23	506	78	17	83	8	465	112	38	63

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
Beginner	2	547	21	0	100	2	571	18	0	100	.	.	.	.	.
Intermediate	.	.	.	.	.	2	535	16	0	100	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	2	556	83	0	100	1	459	.	100	0	.	.	.	.	.
Title III First Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	1	574	.	0	100	.	.	.	.	.
English Speaker I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English Speaker II	1212	490	73	32	68	1058	509	82	32	68	405	511	80	34	66
Pre-functional Waiver	.	.	.	.	.	2	508	54	50	50	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Unknown	1	614	.	0	100	3	467	13	67	33	.	.	.	.	.
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	373	473	69	41	59	342	493	74	39	61	115	502	84	39	61
Free	765	500	74	27	73	666	518	84	28	72	257	513	82	32	68
Reduced	114	481	65	40	60	82	504	75	34	66	40	514	67	30	70
Blank	.	.	.	.	.	2	471	16	50	50	1	540	.	0	100
IEP															
Severe Mental Disability	157	396	63	94	6	120	403	67	92	8	48	390	73	96	4
Moderate Mental Disability	338	482	46	30	70	446	501	53	30	70	190	513	51	29	71
Mild Mental Disability	350	541	55	6	94	239	584	66	3	97	87	589	48	3	97
Autism	272	486	61	33	67	189	496	61	37	63	57	481	78	47	53
Deaf/Blindness	5	378	45	100	0	5	473	68	60	40	3	426	159	67	33
Emotional Disability	7	554	67	14	86	5	584	34	0	100	1	592	.	0	100
Hearing Impaired	21	457	65	48	52	20	525	73	20	80	13	530	69	31	69
Learning Disability	6	568	42	0	100	10	608	70	10	90	1	496	.	0	100
Multiple-Disability	104	437	76	64	36	90	457	85	60	40	37	471	82	54	46
Other Health Impaired	54	502	88	30	70	45	512	101	22	78	11	498	102	27	73
Orthopedically Impaired	62	488	71	40	60	46	503	72	33	67	14	502	82	21	79
Speech or Language Impaired	925	495	65	29	71	616	508	70	31	69	141	506	62	37	63
Traumatic Brain Injury	4	493	94	25	75	4	492	29	50	50	5	504	70	40	60
Visually Impaired	49	429	71	76	24	32	455	97	66	34	11	434	98	45	55
TOTAL	1252	490	73	32	68	1092	509	81	32	68	413	510	81	34	66

Exhibit 24: Performance by Grade-Band Form and Student Age—Mathematics

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	139	486	68	45	55	.	.	.	.	.	.	.	.	.	.
9	420	488	72	39	61	.	.	.	.	.	.	.	.	.	.
10	417	499	73	37	63	1	544	.	0	100	.	.	.	.	.
11	4	475	83	25	75	406	506	81	40	60	.	.	.	.	.
12	.	.	.	.	.	348	502	83	47	53	.	.	.	.	.
13	.	.	.	.	.	332	513	83	34	66	.	.	.	.	.
14	.	.	.	.	.	1	619	.	0	100	2	473	42	50	50
15	.	.	.	.	.	.	.	.	.	.	394	498	72	45	55
16	.	.	.	.	.	.	.	.	.	.	16	505	69	31	69
Blank	268	477	70	47	53	1	414	.	100	0	.	.	.	.	.
STUDENT'S ETHNICITY															
African American	637	493	73	38	62	556	514	83	38	62	211	507	69	43	57
African American/American Indian	5	484	58	40	60	1	340	.	100	0	2	543	56	0	100
American Indian	3	470	27	67	33	5	467	113	40	60	.	.	.	.	.
Asian	14	487	47	36	64	10	505	83	40	60	3	415	116	67	33
Hawaiian/Pacific Islander	1	354	.	100	0	2	547	25	0	100	.	.	.	.	.
Hispanic	44	470	65	55	45	40	500	71	38	63	7	488	95	43	57
Other	5	463	47	40	60	2	580	177	50	50	1	497	.	100	0
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
White	518	487	71	42	58	458	500	81	43	57	187	490	72	46	54
White/African American	15	501	101	40	60	11	508	104	27	73	1	493	.	100	0
White/American Indian	1	698	.	0	100	2	511	76	50	50	.	.	.	.	.
White/Asian	5	450	56	80	20	2	495	45	50	50	.	.	.	.	.
STUDENT'S GENDER															
Female	407	479	78	45	55	353	506	86	40	60	153	499	68	46	54
Male	841	494	68	38	62	736	508	81	41	59	259	498	74	43	57
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															
Pre-functional	35	463	56	57	43	22	514	76	27	73	8	450	100	63	38
Beginner	2	567	39	0	100	2	547	34	0	100	.	.	.	.	.

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
Intermediate	.	.	.	.	.	2	532	46	0	100	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	2	562	71	0	100	1	440	.	100	0	.	.	.	.	.
Title III Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	1	568	.	0	100	.	.	.	.	.
English speaker I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English speaker II	1208	490	72	40	60	1056	507	83	41	59	404	499	71	44	56
Pre-functional Waiver	.	.	.	.	.	2	483	33	50	50	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	1	550	.	0	100	3	441	25	100	0	.	.	.	.	.
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	372	472	71	51	49	342	489	74	50	50	116	489	77	45	55
Free	762	498	73	36	64	664	517	85	35	65	255	502	71	46	54
Reduced	114	486	59	41	59	81	507	83	42	58	40	504	57	33	68
Blank	.	.	.	.	.	2	454	13	100	0	1	505	.	0	100
IEP															
Severe Mental Disability	154	397	63	95	5	120	401	63	96	4	49	391	64	100	0
Moderate Mental Disability	338	479	50	47	53	445	498	53	43	57	190	501	45	48	52
Mild Mental Disability	349	539	55	11	89	239	582	68	5	95	86	564	48	6	94
Autism	272	487	56	42	58	189	497	70	45	55	56	476	65	48	52
Deaf/Blindness	5	378	60	100	0	5	531	111	40	60	3	466	54	67	33
Emotional Disability	7	535	88	14	86	5	581	33	0	100	1	564	.	0	100
Hearing Impaired	21	460	67	52	48	20	534	90	30	70	13	528	93	31	69
Learning Disability	6	579	63	0	100	10	614	62	10	90	1	485	.	100	0
Multiple-Disability	104	435	71	71	29	89	460	95	69	31	37	466	87	68	32
Other Health Impaired	54	504	83	30	70	44	514	98	32	68	11	494	70	36	64
Orthopedically Impaired	62	478	58	47	53	45	496	70	47	53	14	495	77	36	64
Speech or Language Impaired	923	495	65	38	62	615	505	73	40	60	140	493	61	49	51
Traumatic Brain Injury	4	483	114	25	75	4	491	35	50	50	5	514	100	60	40
Visually Impaired	49	431	68	80	20	32	454	104	72	28	11	443	84	73	27
TOTAL	1248	489	72	41	59	1089	507	82	40	60	412	498	72	44	56

Exhibit 25: Performance by Grade-Band Form and Student Age—Science

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	80	494	66	35	65	.	.	.	.	.	.	.	.	.	.
9	420	488	69	30	70	.	.	.	.	.	.	.	.	.	.
10	232	496	77	29	71	1	522	.	0	100	.	.	.	.	.
11	2	496	4	0	100	213	509	76	35	65	.	.	.	.	.
12	.	.	.	.	.	344	501	82	39	61	.	.	.	.	.
13	.	.	.	.	.	190	515	81	31	69	1	328	.	100	0
14	.	.	.	.	.	1	611	.	0	100	2	463	74	50	50
15	.	.	.	.	.	.	.	.	.	.	393	502	77	49	51
16	.	.	.	.	.	.	.	.	.	.	16	506	80	31	69
Blank	131	473	74	44	56	1	441	.	100	0	.	.	.	.	.
STUDENT'S ETHNICITY															
African American	449	491	73	30	70	384	515	81	32	68	211	509	77	44	56
African American/American Indian	5	497	48	40	60	1	404	.	100	0	2	502	27	50	50
American Indian	2	482	21	50	50	2	503	133	50	50	.	.	.	.	.
Asian	11	488	41	45	55	9	490	71	67	33	3	439	126	67	33
Hawaiian/Pacific Islander	1	301	.	100	0	2	533	21	0	100	.	.	.	.	.
Hispanic	31	473	68	42	58	27	493	76	41	59	7	481	106	57	43
Other	2	505	21	0	100	1	508	.	0	100	1	506	.	0	100
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
White	349	486	70	34	66	313	499	78	39	61	187	494	77	52	48
White/African American	11	484	93	36	64	8	514	132	50	50	1	483	.	100	0
White/American Indian	1	733	.	0	100	2	521	92	50	50	.	.	.	.	.
White/Asian	3	471	40	33	67	1	498	.	0	100	.	.	.	.	.
STUDENT'S GENDER															
Female	286	478	77	37	63	246	502	81	38	62	153	501	73	50	50
Male	579	494	69	30	70	504	509	80	35	65	259	501	80	47	53
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															
Pre-functional	24	479	67	38	63	17	500	81	29	71	8	442	101	75	25

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
Beginner	1	527	.	0	100	2	568	25	0	100	.	.	.	.	.
Intermediate	.	.	.	.	.	2	556	8	0	100	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	1	511	.	0	100	1	472	.	100	0	.	.	.	.	.
Title III Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	1	554	.	0	100	.	.	.	.	.
English speaker I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English speaker II	838	489	72	32	68	725	507	80	36	64	404	502	77	48	52
Pre-functional Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	1	590	.	0	100	2	470	40	50	50	.	.	.	.	.
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	255	472	69	42	58	223	484	73	48	52	116	490	77	57	43
Free	533	497	74	28	72	462	519	81	29	71	255	505	81	45	55
Reduced	77	489	59	31	69	64	502	82	39	61	40	506	60	43	58
Blank	.	.	.	.	.	1	498	.	0	100	1	495	.	100	0
IEP															
Severe Mental Disability	107	394	67	91	9	81	390	70	98	2	49	389	67	100	0
Moderate Mental Disability	221	483	56	33	67	315	503	53	35	65	189	508	50	50	50
Mild Mental Disability	249	535	47	6	94	170	576	56	3	97	86	572	47	6	94
Autism	187	485	56	36	64	123	496	64	46	54	56	470	74	66	34
Deaf/Blindness	4	359	84	100	0	3	509	70	33	67	3	429	59	100	0
Emotional Disability	7	558	85	0	100	3	572	35	0	100	1	620	.	0	100
Hearing Impaired	16	466	66	31	69	14	540	91	29	71	13	518	88	38	62
Learning Disability	5	565	10	0	100	8	611	79	13	88	1	491	.	100	0
Multiple-Disability	71	435	72	66	34	67	455	96	67	33	37	468	88	70	30
Other Health Impaired	40	508	87	25	75	28	527	97	11	89	11	490	94	36	64
Orthopedically Impaired	48	475	69	40	60	29	489	67	52	48	14	491	89	50	50
Speech or Language Impaired	641	496	64	29	71	419	507	73	36	64	141	494	64	57	43
Traumatic Brain Injury	4	468	118	25	75	2	506	30	50	50	6	480	95	50	50
Visually Impaired	35	426	79	71	29	22	445	109	77	23	11	442	89	64	36
TOTAL	865	489	72	32	68	750	507	80	36	64	412	501	78	48	52

Exhibit 26: Performance by Grade-Band Form and Student Age—Social Studies

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
STUDENT'S AGE															
8	70	486	67	54	46	.	.	.	.	.	.	.	.	.	.
9	416	496	74	46	54	.	.	.	.	.	.	.	.	.	.
10	210	513	80	38	62	.	.	.	.	.	.	.	.	.	.
11	3	458	86	67	33	214	514	100	44	56	.	.	.	.	.
12	.	.	.	.	.	340	513	88	42	58	.	.	.	.	.
13	1	260	.	100	0	184	524	86	37	63	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	152	485	71	52	48	.	.	.	.	.	.	.	.	.	.
STUDENT'S ETHNICITY															
African American	419	505	74	42	58	374	521	94	39	61	.	.	.	.	.
African American/American Indian	3	501	48	33	67	.	.	.	.	.	.	.	.	.	.
American Indian	3	490	21	33	67	4	516	80	25	75	.	.	.	.	.
Asian	7	486	69	57	43	9	497	81	56	44	.	.	.	.	.
Hawaiian/Pacific Islander	1	321	.	100	0	2	545	22	0	100	.	.	.	.	.
Hispanic	32	492	66	47	53	27	499	87	44	56	.	.	.	.	.
Other	4	437	51	75	25	2	621	158	0	100	.	.	.	.	.
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
White	367	491	78	49	51	312	511	87	45	55	.	.	.	.	.
White/African American	11	468	87	55	45	6	526	142	33	67	.	.	.	.	.
White/American Indian	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
White/Asian	5	472	47	60	40	2	483	14	100	0	.	.	.	.	.
STUDENT'S GENDER															
Female	281	486	77	48	52	248	514	95	40	60	.	.	.	.	.
Male	571	503	74	44	56	490	517	89	42	58	.	.	.	.	.
Blank	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (LANGUAGE)															
Pre-functional	24	480	58	54	46	15	523	83	27	73	.	.	.	.	.
Beginner	2	571	5	0	100	2	588	27	0	100	.	.	.	.	.

	Elementary					Middle					High				
		Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)			Scale Score		Ach. Level (%)	
	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3	N	Mean	SD	< 3	>= 3
Intermediate	.	.	.	.	.	1	540	.	0	100	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Initially English Proficient	2	563	54	0	100	.	.	.	.	.	.	.	.	.	.
Title III Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	1	534	.	0	100	.	.	.	.	.
English speaker I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English speaker II	823	497	76	46	54	716	516	92	42	58	.	.	.	.	.
Pre-functional Waiver	.	.	.	.	.	2	500	41	50	50	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blank	1	630	.	0	100	1	493	.	100	0	.	.	.	.	.
ELIGIBLE FOR FREE OR REDUCED LUNCH															
No	253	477	77	59	41	236	495	83	53	47	.	.	.	.	.
Free	516	509	73	38	62	441	527	94	36	64	.	.	.	.	.
Reduced	83	487	71	52	48	60	518	83	40	60	.	.	.	.	.
Blank	.	.	.	.	.	1	493	.	100	0	.	.	.	.	.
IEP															
Severe Mental Disability	103	397	68	98	2	88	403	73	97	3	.	.	.	.	.
Moderate Mental Disability	237	492	59	55	45	308	515	63	40	60	.	.	.	.	.
Mild Mental Disability	245	546	55	13	87	154	598	70	5	95	.	.	.	.	.
Autism	182	493	61	51	49	117	489	70	58	42	.	.	.	.	.
Deaf/Blindness	4	340	62	100	0	5	500	70	40	60	.	.	.	.	.
Emotional Disability	5	575	119	20	80	3	627	70	0	100	.	.	.	.	.
Hearing Impaired	14	456	72	64	36	16	525	91	38	63	.	.	.	.	.
Learning Disability	4	575	9	0	100	8	624	79	13	88	.	.	.	.	.
Multiple-Disability	70	443	82	76	24	55	462	89	71	29	.	.	.	.	.
Other Health Impaired	31	499	76	29	71	30	508	114	33	67	.	.	.	.	.
Orthopedically Impaired	41	478	77	59	41	35	528	82	34	66	.	.	.	.	.
Speech or Language Impaired	626	503	66	43	57	411	509	80	45	55	.	.	.	.	.
Traumatic Brain Injury	2	540	13	0	100	3	519	48	33	67	.	.	.	.	.
Visually Impaired	29	422	79	90	10	20	457	107	75	25	.	.	.	.	.
TOTAL	852	497	76	46	54	738	516	91	41	59	.	.	.	.	.

Chapter 8: Validity

Content Validity

One source of evidence for the content validity of the South Carolina Alternate Assessment was obtained through independent alignment studies. The University of North Carolina at Charlotte (UNCC) conducted studies of the alignment of (a) ASMGs to grade-level curriculum standards and (b) SC-Alt items to the ASMGs that they targeted. This was a pilot study conducted by Flowers, Browder, Wakeman, and Karvonen with UNCC through the National Alternate Assessment Center (NAAC). (South Carolina is a member state of the NAAC.) A second independent study of ELA and mathematics was completed by the South Carolina Education Oversight Committee (EOC; 2008a) as required by the state Education Accountability Act of 1998 (EAA). The EOC approved the ELA and mathematics content areas on February 28, 2008. The UNCC alignment study results for the English language arts and mathematics assessments are reported in detail in Flowers, Browder, Wakeman, and Karvonen (2006a). The results of the alignment studies for the ELA and mathematics assessments indicate that

> The state has evidence supporting alignment for its measurement guidelines and alternate assessment based on all seven criteria. We conclude that overall this is an alternate assessment system that links to the grade level content. Some areas for consideration in further development of the system are noted related to balance of content. (p. 7)

The alignment study results for the science assessment are reported in detail in Flowers, Browder, Wakeman, and Karvonen (2006b) and in an addendum dated December 21, 2007. The results of the alignment study for the science assessment indicate that

> The strength of the South Carolina science Alternate Assessment was that nearly all of the content was academic science content (98%). This is especially notable given that the alternate assessment tasks included items accessible to students at all symbolic levels. In contrast, the degree of alignment of AA tasks/items to grade-level standards was lower than those found in the alignment of ELA and mathematics. This difference could be due to the fact that the state's science grade-level standards changed during the development of the science AA. Another challenge was that the state had linked its alternate assessment tasks to the state standards and not directly to the measurement guidelines, creating a tough challenge to demonstrating alignment…. Our work with other states suggests that science may typically be the area rated as having the weakest alignment. (p. 4).

SCDE reviewed the initial science alignment study and determined that one source of some misalignment had resulted from the linking of some items to multiple standards and indicators in the alignment document provided by SCDE. During the Science Content Review Committee meeting, some members recommended adding additional indicators to align to some items. The intent of these recommendations focused more on instruction and demonstrating that instruction could include multiple standards and indicators. However, the alignment study team considered only the first two standards aligned to each item. In some cases, the first two

standards were not necessarily the most appropriate. SCDE prioritized the standards and indicators and resubmitted the documentation for an additional study. From this review, completed December 21, 2007 (Flowers, Browder, Wakeman, and Karvonen, 2007), 163 of 173 items were rated as academic. Of the 10 items listed as nonacademic, six were rated as foundational (p.1). SCDE is currently addressing the items that were rated as having no content centrality by developing replacement items for new forms.

At the time of the alignment study for ELA and mathematics by Dr. Flowers and colleagues, the design of the SC-Alt was envisioned as a single assessment across grade levels. This design changed to a grade-band assessment following the study; however, the information provided from the alignment study was used to identify items with alignment difficulty, and these items were omitted from the operational grade-band test forms. Information from the review along with teacher comments was also used during item data review as part of the decision-making process regarding inclusion of items in the assessment.

A second independent review of the alignment of the science assessment was conducted by the Education Oversight Committee (EOC; 2008b) The EOC approved the elementary and middle school science alternate assessment on August 12, 2008. The EOC alignment findings were based on the review of two sets of studies of the SC-Alt:

- studies of the alignment between the SC-Alt Science assessment and the state academic standards conducted by University of North Carolina-Charlotte and Western Carolina University professors of curriculum and special education, in cooperation with the South Carolina State Department of Education (SCDE) and the National Alternate Assessment Center (Flowers, Browder, Wakeman, & Karvonen, December 2006; January 2007; December 2007); and
- a technical review of the task and item data from the 2007 test administration conducted by a professor of educational research and assessment at the University of South Carolina.

Copies of the reports of the EOC reviews and findings are available in their entirety from the SCDE. Based on this review, the EOC identified a number of strengths of the SC-Alt science assessment that were noted in the final report:

- The assessment provides accountability and information for instructional improvement for students with significant cognitive disabilities who would not otherwise be assessed in the state testing programs, even with test accommodations and modifications.
- The assessment is intended to be aligned with the same grade level academic standards as for all students, although at levels of complexity appropriate for the diversity of cognitive functioning observed among students with significant cognitive disabilities.
- The assessment format allows students to respond to the items using the communication modes the student uses during instruction, such as oral response, pointing, use of eye gaze, use of a response card, sign language, or an augmentative communication device.
- The procedures for placing the student at the appropriate level for beginning each assessment reduces student fatigue and maximizes students' opportunities to show their highest performance;
- The items in the assessment have a wide range of difficulty and the test is moderately able to discriminate between high and low levels of performance.

The EOC report noted that while 96% of the items were found to be aligned to science inquiry standard indicators, the alignment of the items to content standards was 78%, falling short of an expectation for successful alignment of 90% set by the original evaluators. The EOC recommended that the SCDE review the alignment of the SC-Alt science items to the grade-level standards and identify items needing revision or replacement.

The SCDE and its contractor, the American Institutes for Research (AIR) reviewed the alignment and the Assessment Standards and Measurement Guidelines (ASMGs) and established priorities for development of tasks to fill identified gaps. During 2008, SCDE and AIR developed five new tasks consisting of 32 items to be used to replace poorly aligned items and improve content coverage in science. Three tasks were developed for the elementary science form and two tasks were developed for the middle school form based on the findings of the alignment study. No new tasks were developed for the high school physical science test as a stand-alone biology field test was planned for Spring 2010. In the future, biology end-of-course tests will replace the physical science end-of-course tests for purposes of reporting ESEA participation calculations.

An independent review of the alignment of the new items by the Center for Research on Education (2009a) found that 98% of the new items were aligned to grade-level content standard indicators. Copies of the report of the alignment reviews and findings are available in their entirety from the SCDE.

A follow up alignment study of the high school ELA and mathematics assessments and biology field test items was conducted by the Center for Research on Education in October 2009 using the same procedures as used for the elementary and middle school alignment studies in December 2006 and January 2007. Almost all (94% to 96%) of the items were rated as academic. This percentage exceeds the value typically found in alternate assessments (i.e. 90%) according to the reviewers. The alignment study results are reported in detail in High School Alternate Assessment Alignment Study Report to the South Carolina Department of Education, October 2009 (Center for Research on Education, 2009b).

Convergent and Discriminant Validity

According to Critical Element 4.1(e) of the federal peer review and Standard 1.14 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), it is desirable, if not necessary, to provide evidence of convergent and discriminant validity. One common method for examining this aspect of validity is with a multitrait-multimethod matrix (MTMM) (Campbell & Fiske, 1959).

Campbell and Fiske (1959) proposed the MTMM matrix design as a tool for the study of convergent and discriminant validity in psychological measurement. The MTMM matrix employs a crossed factorial measurement design of traits and methods to reveal these types of validity in comparison:

- Large correlations on validity diagonals (i.e., same trait and different methods) indicate convergent validity.
- Low correlations in the heterotrait-monomethod blocks indicate discriminant validity and the absence of method effects.

- Low correlations in the heterotrait-heteromethod blocks also indicate discriminant validity.

Selection of Traits and Methods

The student's abilities in each of the subjects—ELA, mathematics, science, and social studies—make up the four traits for the MTMM study. Two methods are considered for assessing these traits: the Student Placement Questionnaire (SPQ) as a structured teacher rating of student ability and the SC-Alt scale score as an IRT score of the student's responses to the set of presented test items. In other words, the two methods contrast test scores of student performance with expert (or teacher) ratings. With four traits and two methods, the MTMM correlation matrix is of order 8. Note that the grade 10 assessment does not include a social studies component; therefore, the MTMM matrix for grade 10 has only six rows and columns.

Results

MTMM matrices were computed separately for each grade band. The results are given in Exhibit 27 through Exhibit 29. Pearson correlations are used, with pairwise deletion of missing data. For each matrix, the minimum pairwise sample size is indicated. *P*-values of individual correlation coefficients are not reported since *all* correlations are significant ($p < 0.05$).

Exhibit 27: MTMM Matrix, Scale Scores with SPQ Scores, Grades 3–5

		IRT Scale Scores				**SPQ Scores**			
	Subject	**ELA**	**Math**	**Science**	**Social Studies**	**ELA**	**Math**	**Science**	**Social Studies**
IRT Scale Scores	ELA	1.0							
	Math	*0.90*	1.0						
	Science	*0.90*	*0.91*	1.0					
	Social Studies	*0.89*	*0.89*	*0.92*	1.0				
SPQ Scores	ELA	**0.76**	0.75	0.74	0.74	1.0			
	Math	0.76	**0.76**	0.74	0.74	*0.92*	1.0		
	Science	0.72	0.70	**0.71**	0.71	*0.88*	*0.89*	1.0	
	Social Studies	0.75	0.75	0.75	**0.76**	*0.90*	*0.89*	*0.92*	1.0
Minimum pairwise N: 437									

Exhibit 28: MTMM Matrix, Scale Scores and SPQ Scores, Grades 6–8

		IRT Scale Scores				SPQ Scores			
	Subject	**ELA**	**Math**	**Science**	**Social Studies**	**ELA**	**Math**	**Science**	**Social Studies**
IRT Scale Scores	ELA	1.0							
	Math	*0.90*	1.0						
	Science	*0.90*	*0.90*	1.0					
	Social Studies	*0.89*	*0.89*	*0.93*	1.0				
SPQ Scores	ELA	**0.78**	0.77	0.78	0.78	1.0			
	Math	0.79	**0.80**	0.79	0.77	*0.92*	1.0		
	Science	0.77	0.77	**0.77**	0.77	*0.86*	*0.91*	1.0	
	Social Studies	0.78	0.77	0.81	**0.79**	*0.90*	*0.91*	*0.91*	1.0
Minimum pairwise N: 361									

Exhibit 29: MTMM Matrix, Scales Scores and SPQ Scores, Grade 10

		IRT Scale Scores			SPQ Scores		
	Subject	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
IRT Scale Scores	ELA	1.0					
	Math	*0.89*	1.0				
	Science	*0.91*	*0.90*	1.0			
SPQ Scores	ELA	**0.79**	0.78	0.75	1.0		
	Math	0.79	**0.81**	0.77	*0.92*	1.0	
	Science	0.76	0.75	**0.73**	*0.88*	*0.91*	1.0
Minimum pairwise N: 364							

In each MTMM table, the convergent validity coefficients (correlations between measurements of the same trait using different methods) are marked in bold. These convergent validity coefficients range from 0.71 to 0.81 and certainly fall into an acceptable range. These high correlations indicate good validity for the SPQ. The above three exhibits indicate that the SPQ and the actual test are essentially measuring the same trait and that the SPQ is a good indicator of performance on the test.

The entries in the mono-method triangles (correlations between measurements of different traits using the same method) are set in italics. These correlation coefficients range between 0.89 and 0.93 for IRT scale scores and between 0.86 and 0.92 for SPQ scores. The high overall range of these correlations indicates the presence of method variance. However, this is to be expected because the SPQ was not developed to measure the trait; instead, it only indicates the starting task on the test for measuring the trait. Such a result of high correlations in the mono-method triangles is not uncommon in MTMM studies (Fiske, 1995) and specific conditions offer themselves as causes for the present scenario: First, the different scale types—number-correct rating scales versus IRT scales of behavioral tests—are in themselves a source of method variation; secondly, the SPQ's "can do" questions draw on the teacher's memory of a student's

possible performance over the long term and are apt to differ in quality and veracity; and thirdly, the IRT scale scores for the three subjects reflect the student's performance in the testing situation and are subject to the student's form on the testing day.

The heterotrait-heteromethod correlations appear in the tables in regular type. These correlation coefficients fall into the same range as the convergent validity coefficients, with values from 0.70 to 0.81. To confirm discriminant validation, the heterotrait-heteromethod correlations should be smaller than the convergent validity coefficients. Instead, these MTMM matrices support the notion that that the four traits vary essentially along just a single dimension. Because the population of alternate assessment students is so *very* heterogeneous, the students' general level of cognitive functioning dominates the relationship among their scale scores.

Validity of the Student Placement Questionnaire (SPQ)

AIR reviewed item data from the 2009 administration regarding the agreement between SPQ recommended start points and the final observed start points. The purpose of the study was to determine the effectiveness of the SPQ in identifying the most appropriate starting task.

Administration of the SC-Alt uses the SPQ as a pre-assessment instrument to determine the most appropriate starting point in the assessment. The SPQ requires the teacher to evaluate the student on 12 to 15 "can do" statements addressing the student's skills and knowledge in each content area on the basis of the teacher's prior instructional knowledge of the student. A total score computed from the teacher's SPQ responses indicates the initial starting task for the assessment.

The instructions for using the SPQ require teachers to adjust the starting point below the SPQ recommended start point when the student is not successful on the first administered task. Alternately, after reviewing the assessment, some teachers may have judged that a student needed to start at a higher level than recommended by the SPQ.

A summary of the results of the agreement between the SPQ recommended start points and the observed start points for each content area and grade-band form is presented in Exhibit 30. These results indicate that the agreement between the SPQ recommended start point and the observed start point was 92% for ELA, 91% for mathematics, 86% for science, and 89% for social studies administrations. Use of the SPQ pre-assessment score is only the first step in the procedure used by the test administrator in determining where the student should start the assessment. Since the test administrator is required to make adjustments based on the student's success on the first task, and these adjustments are reflected in the agreement rates, the SPQ appears to be working very effectively for targeting the first task to begin the assessment process.

Exhibit 30: Agreement Between SPQ and Observed Start Points by SPQ Recommended Starting Tasks

		Elementary School				Middle School				High School				
		Recommended Starting Task												
Subject	**Observed Start Task**	**1**	**3**	**6**	**Total**	**1**	**3**	**6**	**Total**	**1**	**3**	**6**	**Total**	**Overall**
ELA	Starting task consistent with SPQ	346	298	516	1160	243	188	581	1012	92	63	213	368	2540
	Lower start task than recommended	0	8	20	28	0	9	6	15	0	7	10	17	60
	Higher start task than recommended	2	4	0	6	2	5	0	7	1	0	0	1	14
	Non-standard start task	8	5	1	14	7	1	4	12	4	1	1	6	32
	Incomplete SPQ	0	0	0	44	0	0	0	46	0	0	0	21	111
	Inconsistent with SPQ (%)	2.81	5.40	3.91	7.35%	3.57	7.39	1.69	7.33%	5.15	11.27	4.91	10.90%	7.87
	ELA Total (N)	356	315	537	1252	252	203	591	1092	97	71	224	413	2757
Math	Starting task consistent with SPQ	315	342	482	1139	239	206	555	1000	87	69	201	357	2496
	Lower start task than recommended	0	21	13	34	0	17	4	21	0	7	13	20	75
	Higher start task than recommended	2	5	0	7	4	0	0	4	1	1	0	2	13
	Non-standard start task	6	1	4	11	6	1	1	8	2	0	2	4	23
	Incomplete SPQ	0	0	0	57	0	0	0	56	0	0	0	29	142
	Inconsistent with SPQ (%)	2.48	7.32	3.41	8.73%	4.02	8.04	0.89	8.17%	3.33	10.39	6.94	13.35%	9.20
	Math Total (N)	323	369	499	1248	249	224	560	1089	90	77	216	412	2749

		Elementary School				Middle School				High School				
		Recommended Starting Task												
Subject	**Observed Start Task**	**1**	**3**	**6**	**Total**	**1**	**3**	**6**	**Total**	**1**	**3**	**6**	**Total**	**Overall**
Science	Starting task consistent with SPQ	270	196	242	708	191	137	352	680	105	72	177	354	1742
	Lower start task than recommended	0	8	65	73	0	2	6	8	0	5	12	17	98
	Higher start task than recommended	9	5	0	14	2	5	0	7	2	1	0	3	24
	Non-standard start task	17	2	3	22	7	2	0	9	1	0	3	4	35
	Incomplete SPQ	0	0	0	48	0	0	0	46	0	0	0	34	128
	Inconsistent with SPQ (%)	8.78	7.11	21.94	18.15%	4.50	6.16	1.68	9.33%	2.78	7.69	7.81	14.08%	14.06
	Science Total (N)	296	211	310	865	200	146	358	750	108	78	192	412	2027
Social Studies	Starting task consistent with SPQ	208	179	387	774	117	99	420	636	x	x	x	x	1410
	Lower start task than recommended	0	17	7	24	0	7	10	17	x	x	x	x	41
	Higher start task than recommended	1	2	0	3	8	2	0	10	x	x	x	x	13
	Non-standard start task	7	1	2	10	26	1	1	28	x	x	x	x	38
	Incomplete SPQ	0	0	0	41	0	0	0	47	x	x	x	x	88
	Inconsistent with SPQ (%)	3.70	10.05	2.27	9.15%	22.52	9.17	2.55	13.82%	x	x	x	x	11.32
	Social Studies Total (N)	216	199	396	852	151	109	431	738	x	x	x	x	1590
	OVERALL	1191	1094	1742	4217	852	682	1940	3669	295	226	632	1237	9123

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC.

American Institutes for Research. (2007). *South Carolina Alternate Assessment Spring 2007 Standard Setting Technical report.* Washington, DC: Author**.**

American Institutes for Research (2008) *South Carolina Alternate Assessment (SC-Alt): Technical report for English Language Arts and Mathematics Field Test Administration, Spring 2006*. Washington, DC: Author

American Institutes for Research and South Carolina Department of Education (2008) *South Carolina's Alternate Assessment, SC-Alt Spring 2007 Operational Administration*. Washington, DC: American Institutes for Research

American Institutes for Research and South Carolina Department of Education (2009) *South Carolina's Alternate Assessment, SC-Alt Spring 2008 Operational Administration*. Washington, DC: American Institutes for Research

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Center for Research on Education (2009a). *Alternate assessment alignment study of replacement items report to the South Carolina State Department of Education*, Columbia, SC: South Carolina State Department of Education.

Center for Research on Education (2009b). *High school alternate assessment alignment report to the South Carolina State Department of Education*, Columbia, SC: South Carolina State Department of Education.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage Publications.

Dorans, N., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23,* 355–368.

Education Oversight Committee (2008a). *Review of the SC-Alternate English language arts and mathematics assessments report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education

Education Oversight Committee (2008b). *Review of the SC-Alternate science assessment report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education

Ferrara, S., Perie, M., & Johnson, E. (2008). Matching the judgmental task with standard setting panelist expertise: The Item-Descriptor (ID) Matching procedure. *Journal of Applied Testing Technology*, 9, 1-22.

Fiske, D. W. (1995) Reprise, new themes, and steps forward. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2006a). *Alternate assessment alignment pilot study report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2006b). *Science alternate assessment alignment pilot study report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2007). *Science alternate assessment alignment pilot study report to the South Carolina State Department of Education, Addendum, December 21, 2007*. Columbia, SC: South Carolina State Department of Education.

Huynh, H. (2003, August). *Technical memorandum for computing standard error in Bookmark standard setting* [The South Carolina PACT 2003 Standard Setting Support Project]. Columbia, SC: University of South Carolina.

Kleinert, H. L., and Kearns, J. F. (2001). *Alternate assessment: Measuring outcomes and supports for students with disabilities.* Baltimore, MD: Brookes Publishers.

Linacre, J. M., & Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370. Available at http://www.rasch.org/rmt/rmt83.htm.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

National Alternate Assessment Center (2005) *Designing from the Ground Floor*. Conference Pre-Session, Chief Council for State School Officers' Large-Scale Assessment Conference. San Antonio, TX: June 17-18, 2005.

Rudner, L.M. & Schafer, W.D. (2001) Reliability (ERIC Digest). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation (ERIC Document Reproduction Service No. ED458213). Retrieved May 17, 2009 from http://www.ericdigests.org/2002-2/reliability.htm

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, *28*(3), 234–247.

Spitzer, R., Cohen, J., Fleiss, J., & Endicott, J. (1967). Quantification of agreement in psychiatry diagnosis: A new approach. *Archives of General Psychiatry*, *17*, 83–87.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, *30*, 233–251.

Zwick, R., & Thayer, D. T. (1996). Evaluating the magnitude of differential item functioning in polytomous items. *Journal of Educational and Behavioral Statistics*, *21*(3), 187–201.

Appendices

Appendix A: Assignment of Tasks to Grade-Band Forms for the Spring 2009 Administration

All tasks in each SC-Alt grade-band assessment align with the Assessment Standards and Measurement Guidelines (ASMGs) in that grade band. Because adjacent grade-band score scales are linked psychometrically, some tasks in each grade-band assessment align with ASMGs in both adjacent grade bands. In turn, these separate grade-band ASMGs link to separate grade-level performance standards for the appropriate grades.

All items in linking tasks are developed to be appropriate for students in both adjacent grade bands. In some cases (e.g., some tasks in ELA), the ASMGs to which linking tasks align are equivalent for two adjacent grade bands. However, the grade-level performance standards to which the ASMGs are linked *do* differ across the adjacent grade bands. In all content areas and for all grade bands, Descriptions of Achievement Levels (DALs) are specific to each grade band and differ across grade bands.

2009 Operational Test Designs to Support Psychometric Linking of Grade-Band Score Scales

To provide data to link all grade-band assessments onto a vertical scale, linking tasks were repeated in adjacent grade assessments. For example, five of the tasks that appeared in the ELA grades 3–5 assessment also appeared in the ELA grades 6–8 assessment. Those five linking tasks and the seven unique tasks made up the 12 tasks in the ELA grades 3–5 assessment. The ELA grades 6–8 assessment included the five linking tasks from the 3–5 assessment, five linking tasks that also appeared in the grades 9–12 assessment, and two unique tasks. This "linking upward" design ensures that students were assessed on ASMGs aligned with their current grade placement or previous grades.

The ELA and mathematics forms (all grade bands) and the science and social studies forms for the elementary and middle school grade bands, were each followed by a set two to three "embedded" field test tasks. Test administrators were instructed to administer these field test tasks to all students, independent of whether the students had taken the low-, intermediate-, or high-complexity sets of items in the subject. The field test items were unique to each form level. Calibrated to the same scale as the operational items, the field test items were linked across grade bands via the linking tasks of the operational test forms.

English Language Arts Assessment

The ELA assessment covered ASMGs in reading, writing, and communication. The previously-designed operational test forms included 12 tasks for each of three grade-band assessments and a total of 26 tasks; 2 or 3 field-test tasks were developed for each grade-band, resulting in a total of 14 or 15 tasks per grade-band and 34 tasks overall. The design for the ELA assessment for spring 2009 appears in Exhibit 31.

Exhibit 31: Numbers of Tasks in Each Operational Grade-Band Assessment, ELA

Grade Band	Unique Tasks	Linking Tasks		Total for Operational Test
	Tasks in Each Grade-Band Assessment Test Booklet			
10	7 (+2 FT tasks)	5	—	14
6–8	2 (+3 FT tasks)		5	15
3–5	7 (+3 FT tasks)	—		15
	Tasks to Be Included			
All grades	24	10		34

Mathematics Assessment

The mathematics assessment covered the mathematics ASMGs. The previously-designed operational test forms included 12 tasks for each of three grade-band assessments for a total of 22 tasks; 2 field-test tasks were developed for each grade-band resulting in 14 tasks per grade-band and 28 tasks overall. The design for the mathematics assessment for spring 2009 appears in Exhibit 32.

Exhibit 32: Numbers of Tasks in Each Operational Grade Band Assessment, Mathematics

Grade Band	Unique Tasks	Linking Tasks			Total for Operational Test
	Tasks in Each Grade-Band Assessment Test Booklet				
10	6 (+2 FT tasks)	2	—	4	14
6–8	2 (+2 FT tasks)		4		14
3–5	4 (+2 FT tasks)	—			14
	Tasks to Be Included				
All grades	18	10			28

Science Assessment

The science assessment covered the science ASMGs. The previously-designed operational test forms included 12 tasks for each of three grade-band assessments and a total of 27 tasks; 3 field-test tasks were added for grade-band 3-5 and 2 field-test tasks were added for grade-band 6-8 for an overall total of 32 tasks. The design for the science assessment for spring 2009 appears in Exhibit 33.

Exhibit 33: Numbers of Tasks in Each Operational Grade-Band Assessment, Science

Grade Band	Unique Tasks	Linking Tasks		Total for Operational Test
	Tasks in Each Grade-Band Assessment Test Booklet			
10	8	4	—	12
6–8	3 (+2 FT tasks)		5	14
3–5	7 (+3 FT tasks)	—		15
	Tasks to Be Included			
All grades	23	9		32

Social Studies Assessment

The social studies assessment covered the social studies ASMGs. The previously-designed test forms included 12 operational for each of two grade-band assessments resulting in a total of 19 tasks; 2 field-test tasks were developed for each grade-band for an overall total of 23 tasks. The design for the science assessment for spring 2009 appears in Exhibit 34.

Exhibit 34: Numbers of Tasks in Each Operational/Field-Test Grade-Band Assessment, Social Studies

Grade Band	**Unique Tasks**	**Linking Tasks**	**Total for Operational Test**
Tasks in Each Grade-Band Assessment Test Booklet			
6–8	7 (+2 FT tasks)	5	14
3–5	7 (+2 FT tasks)		14
Tasks to Be Included			
All grades	18	5	23

Appendix B: Starting and Stopping Rules for Using the Student Placement Questionnaire

Directions for Determining the Starting and Concluding Tasks and Use of the Student Placement Questionnaire, Spring 2009

These directions guide you through the following steps:

- completing the Student Placement Questionnaire (SPQ),
- identifying the starting task in each content area,
- adjusting the starting task, if that becomes necessary,
- determining when to conclude the administration

Completing the Student Placement Questionnaire

The SPQ is designed to identify the most appropriate starting task for each of your students in each content area of SC-Alt. You will use the SPQ to identify the most appropriate starting task for each student in the SC-Alt assessments in English language arts, mathematics, science, and social studies. Answer each SPQ item as accurately as you can based on your experience in the classroom with this student.

The SPQs are located in the Student Answer Folder along with the areas for recording the student's scores on each SC-Alt task. An example of a completed English Language Arts SPQ is included at the end of these instructions.

Identifying the Starting Task for a Student in Each Content Area

1. Bubble in your responses to the SPQ questions.

   After you respond to all items in the SPQ, identify the most appropriate starting task for this student following the steps on the SPQ. These are the steps:

2. Count the number of bubbles you marked in each of the first three columns, and write the totals in the blocks under each column.
3. In section 3 at the bottom of the page:
   a. Write the column totals in the appropriate blocks.
   b. Multiply each total by the specified multiplier, and write the resulting totals in the blocks to the right.
   c. Sum the three totals to obtain the total SPQ score. Write the SPQ score into the blocks and bubble in the SPQ score.

   Please check your work and complete the bubble grids for the total SPQ score.
4. Find the total SPQ score in section 4 to determine the starting task for this student.

Administering the Starting Task and Completing the Administration

After you identify the starting task for this student using the SPQ, follow these directions to administer the starting task and complete the administration.

The SPQ provides the initial starting point for a student's administration. Each student must be administered a minimum of five tasks (including the starting task) if the student is started at Task 1 or a minimum of seven tasks if the student is started at Task 3 or Task 6. The minimum number of tasks and specific tasks that must be administered to each student for each starting level are specified in the table below.

Exhibit 35: Minimum Task Ranges to Be Administered

ELA, Mathematics, Science, and Social Studies	
Starting task	**Administer all items in at least these tasks**
Task 1	1–5
Task 3	3–9
Task 6	6–12

It may be necessary to adjust the starting task based on the student's level of success on the first task. Also, the administration should be continued beyond the minimum number of tasks when the student is responding successfully.

When the Student Does Not Respond Successfully on the First Task

"Responding successfully" means getting at least three total points on a task. Each task has at least four items. Responding successfully would mean that a student received at least three total points for all the items combined. For example, a student may respond successfully by receiving three points on one item, or two points on one item and one point on another item, or one point each on three different items. When a student does not receive three or more total points on a task, the student has not responded successfully on the task.

When a student is started at Task 3 or at Task 6 and does not respond successfully on the first task, the starting task was too difficult, and the teacher must restart the student at the next lower starting point. For example:

- If the student starts at Task 3 but cannot respond successfully on Task 3, restart the student at Task 1.
- If the student starts at Task 6 but cannot respond successfully on Task 6, restart the student at Task 3.

When a student is started at Task 1, no downward adjustment is possible, and the administration must progress through at least five tasks.

When to Conclude the Administration

If the student responds successfully on the last required task as specified in the table above, continue with the administration by administering the next task and subsequent tasks until the student no longer responds successfully on a task. By continuing the administration of

subsequent tasks when the student is "responding successfully," you will provide the maximum opportunity for the student to demonstrate his or her knowledge and skills.

If the student does not respond successfully on the last required task or if at any point the student does not respond successfully on additional tasks (i.e., obtain three or more points on the task), you may conclude the administration.

If you conclude the administration after administering the required tasks and when the student is no longer successful, you will not prolong the student's test administration unnecessarily and you will avoid any possible negative effects on the student.

Examples:

- Student A was started at Task 1 and administered Tasks 1–5. The student responded successfully on Task 5 and therefore was administered Task 6. The student responded successfully on Task 6 and was administered Task 7. The student did not respond successfully on Task 7, and the administration was concluded after Task 7.
- Student B was started at Task 3 and was administered Tasks 3–9. The student did not respond successfully on Task 9, and the administration was concluded after Task 9.

Additional Special Instructions for the Field Test Administrations: There are additional field test tasks appended to the end of each test booklet, except at grade 10 science. These field test tasks are to be administered to all students who take the assessment. When the student has completed the concluding task, the test administrator should turn to the field test tasks (Tasks 13 – 14 for all mathematics levels, middle school science, all social studies levels, and high school ELA, and Tasks 13 – 15 for elementary and middle school ELA and elementary school science) and administer these tasks.

SC - ALT STUDENT PLACEMENT QUESTIONNAIRE
ENGLISH LANGUAGE ARTS

(completed SPQ example)

Follow steps 1-4 to complete the SPQ and identify the starting task.

(1) Please darken the bubble (●) that corresponds to the most appropriate response for this student. Mark only one response for each item. Please mark a response for all items below. Use a No. 2 pencil only.

	Independently	With verbal/gestural prompting	With physical prompting/hand-over-hand	No, she/he cannot do this
In reading, can this student:				
1. Attend to text read aloud?	○	●	○	○
2. Recall details of text read aloud?	○	●	○	○
3. Recognize some high-frequency written words?	○	○	●	○
4. Draw conclusions or make inferences about texts?	○	○	○	●
In writing, can this student:				
5. Write his or her name using a pencil, name stamp, letter titles, or other means?	○	●	○	○
6. Use objects, pictures, and/or picture symbols to write in any format?	○	○	●	○
7. Copy, trace, or print letters?	○	●	○	○
8. Use oral language and/or letters and words to write?	○	○	○	●
In communicating, can this student:				
9. Listen (i.e., demonstrate receptive behavior) and respond?	●	○	○	○
10. Participate in conversations by responding appropriately?	○	●	○	○
11. Use language to express a preference, opinion, or viewpoint?	○	●	○	○
12. Recognize and understand the meaning of environmental signs (e.g., street signs, store signs, school signs)?	○	●	○	○
(2) Write in the total number of bubbles you marked in each column	1	7	2	
	Col. 1	*Col. 2*	*Col. 3*	

(3) Calculate the SPQ total score

(a) write the column totals from (2) in (a) below
(b) multiply and write the results in (b) below
(c) sum the results from (b) and write the sum in (c)

	(a)		(b)
Column 1 Total	1	x 3 =	3
		+	
Column 2 Total	7	x 2 =	14
		+	
Column 3 Total	2	x 1 =	2

(c) Total SPQ Score = 19

(4) Identify the starting task for this student using the SPQ total score from step (3).

If the total score is in this range	Start at this task	Administer all items in at least these tasks
0-11	Task 1	1-5
12-22	Task 3	3-9
23-36	Task 6	6-12

DO NOT WRITE IN THIS AREA

Appendix C: Scoring Audits and Analysis of Video-Rater Data from the

Spring 2009 Operational Administration

A videotaping study was conducted to audit scoring accuracy for the spring 2009 administrations of the SC-Alt in ELA, mathematics, science, and social studies. **Scoring accuracy** refers to the degree to which teachers follow scaffolding and scoring directions correctly and assign correct scores to student responses. This appendix describes the sampling procedures, the identified sample of students, and the results for the attained sample of completed videotaped administrations.

Sampling Procedures

The sampling procedure was designed to include administrations from every school district and to be broadly representative of the range of student and test administrations. A sample of students was identified for videotaping so that (a) all districts implementing the SC-Alt would be required to videotape at least one student administration (all content areas) and (b) the total number of taped administrations per district would be based on the number of teachers involved in the assessment for each district. The sampling was by teacher and student within districts. One-third of the teachers within each district was randomly sampled to videotape the administration of one student. The number of teachers (and students) to be selected from each district is shown in Exhibit 36.

Exhibit 36: District-Based Sampling Targets for Video Rater Study

Total Number of Teachers per District	Number Required to Videotape
1–5	1
6–8	2
9–11	3
12–14	4
15–17	5
18–20	6
21–23	7
24–26	8
27–29	9
30–32	10
33–35	11
36–38	12
39–41	13
42–44	14
45 - 47	15
48 - 50	16
51 - 53	17
54 - 56	18
57 - 59	19
60 - 62	20

Based on this sampling plan and the numbers of pre-identified students coded for each district for the 2009 administration, the frequency distribution of test administrations sampled per district was as follows:

- 1 test administration – 32 districts

- 2-5 test administrations – 26 districts
- 6-10 test administrations – 7 districts
- 11-15 test administrations – 5 districts
- 16-20 test administrations – 1 district

The sampling of students and teachers was conducted from the January 2009 precode file, which was the pre-identification file for the spring 2009 SC-Alt administration. The sampling was conducted by SCDE, and the students identified for videotaping were flagged on the precode file sent to Measurement Incorporated (MI) for the production of materials and district notification. The numbers of students by form and disability sampled for videotaping are reported in Exhibit 37.

Exhibit 37: Stratified Sample of Students Identified for Videotaped Administrations

	Elementary		Middle		High School		Total	
	N	%	N	%	N	%	N	%
PRIMARY DISABILITY								
Severe Mental Disability	16	14.3	9	11.4	7	10.9	32	12.5
Moderate Mental Disability	29	25.9	26	32.9	25	39.1	80	31.4
Mild Mental Disability	34	30.4	18	22.8	18	28.1	70	27.5
Autism	21	18.8	8	10.1	6	9.4	35	13.7
Other Disabilities	12	10.7	18	22.8	8	12.5	38	14.9
TOTAL	112		79		64		255	

Videotaping Procedures

The district test coordinators for alternate assessment were provided rosters of the students identified for videotaping. The district materials included a packet of information for each teacher that included the following:

- a videotaping student roster identifying the student,
- information on the purpose of the videotaping and instructions for how to conduct the videotaped administrations,
- a videotaping student information form,
- bar code labels for positive identification and linking of the videotapes and the student information to the SC-Alt assessment data file, and
- directions for the packaging and return of materials.

Communications to teachers and district test coordinators emphasized the importance of completing the videotaped administrations, provided contact information for questions or concerns, and asked for notification of SCDE if there were a problem in completing a videotaped administration for a particular student. Districts notified SCDE about a small number of students who either could not be assessed (e.g., because the students had moved, the parents did not consent to videotaping, or the students were not going to be assessed with the SC-Alt) or for whom the videotaping was inappropriate or extremely difficult to implement (e.g., medical homebound students or incompatible student behavior due to taping). These students were deleted from the videotaping sample.

Approximately one-half of the students who were deleted from the videotaping sample list by SCDE were replaced by another student with the same teacher or in a few cases by identifying a different teacher and student. The replacement students were selected to match the grade-span form and disability of the original students as closely as possible. As a result of notifications by districts, the SCDE deleted 35 students from the original sample and instructed districts to videotape 17 replacement students.

Analysis of Video Rater Data

The total number of students identified for videotaping after SCDE adjustments (resulting from deletions and replacements) was 237. Videotaping materials were received for 225 of the students. Of these, twelve sets of videotapes were excluded from the analysis due to non-viewable administrations. Additionally, two videotape records could not be linked to operational data. The final number of students in the attained sample was 211. This sample is summarized in Exhibit 38.

Exhibit 38: Demographic Frequencies for the Video Rater Data Sample—by Test Form

	Elementary		Middle		High	
	N	%	N	%	N	%
STUDENT'S ETHNICITY						
African American	49	48.51	30	49.18	29	59.18
African American/American Indian	1	0.99	.	0.00	2	4.08
American Indian	.	0.00	.	0.00	.	0.00
Asian	1	0.99	1	1.64	.	0.00
Hawaiian/Pacific Islander	.	0.00	.	0.00	.	0.00
Hispanic	4	3.96	2	3.28	.	0.00
White	44	43.56	27	44.26	18	36.73
White/African American	2	1.98	.	0.00	.	0.00
White/American Indian	.	0.00	.	0.00	.	0.00
White/Asian	.	0.00	.	0.00	.	0.00
Other	.	0.00	1	1.64	.	0.00
STUDENT'S GENDER						
Female	37	36.63	16	26.23	19	38.78
Male	64	63.37	45	73.77	30	61.22
ESL (LANGUAGE)						
Advanced	.	0.00	.	0.00	.	0.00
Advanced Waiver	.	0.00	.	0.00	.	0.00
Beginner	.	0.00	.	0.00	.	0.00
Beginner Waiver	.	0.00	.	0.00	.	0.00
English Speaker I	.	0.00	.	0.00	.	0.00
English Speaker II	97	96.04	59	96.72	49	100.00
Full English Proficient	.	0.00	.	0.00	.	0.00
Intermediate	.	0.00	.	0.00	.	0.00
Intermediate Waiver	.	0.00	.	0.00	.	0.00
Pre-functional	4	3.96	1	1.64	.	0.00
Pre-functional Waiver	.	0.00	1	1.64	.	0.00
Title III First Year Exited	.	0.00	.	0.00	.	0.00
Title III Second+ Year Exited	.	0.00	.	0.00	.	0.00
ELIGIBLE FOR FREE OR REDUCED-PRICE LUNCH						
Free	66	65.35	38	62.30	29	59.18

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
Reduced	7	6.93	1	1.64	5	10.20
No	28	27.72	22	36.07	15	30.61
EFA GRADE (REPORTED GRADE FOR FUNDING)						
1	1	0.99	.	0.00	.	0.00
2	5	4.95	.	0.00	.	0.00
3	45	44.55	.	0.00	.	0.00
4	30	29.70	1	1.64	.	0.00
5	19	18.81	14	22.95	.	0.00
6	.	0.00	16	26.23	1	2.04
7	1	0.99	14	22.95	1	2.04
8	.	0.00	11	18.03	2	4.08
9	.	0.00	2	3.28	22	44.90
10	.	0.00	3	4.92	23	46.94
11	.	0.00	.	0.00	.	0.00
12	.	0.00	.	0.00	.	0.00
Blank	.	0.00	.	0.00	.	0.00
COMPLETION STATUS: Student satisfied attemptedness rule						
ELA	100	99.01	61	100.00	48	97.96
Mathematics	101	100.00	61	100.00	47	95.92
Science[4]	67	66.34	46	75.41	47	95.92
Social Studies[6]	69	68.32	31	50.82	.	0.00
COMPLETION STATUS: Student did not answer any content area-items						
ELA	1	0.99	.	0.00	1	2.04
Mathematics	.	0.00	.	0.00	2	4.08
Science[6]	33	32.67	15	24.59	2	4.08
Social Studies[6]	31	30.69	30	49.18	49	100.00
COMPLETION STATUS: Student received fewer than five scored responses						
ELA	.	0.00	.	0.00	.	0.00
Mathematics	.	0.00	.	0.00	.	0.00
Science[6]	1	0.99	.	0.00	.	0.00
Social Studies[6]	1	0.99	.	0.00	.	0.00
Migrant Status	.	0.00	.	0.00	.	0.00
Home schooled	.	0.00	.	0.00	.	0.00
Medical Homebound	1	0.99	.	0.00	1	2.04
IEP DISABILITY CODES (Multiple codes per student)						
Severe Mental Disability	14	13.86	6	9.84	6	12.24
Moderate Mental Disability	26	25.74	22	36.07	18	36.73
Mild Mental Disability	31	30.69	17	27.87	15	30.61
Autism	21	20.79	8	13.11	3	6.12
Deaf/Blindness	.	0.00	.	0.00	.	0.00
Emotional Disability	.	0.00	.	0.00	.	0.00
Hearing Impairment	2	1.98	1	1.64	3	6.12
Learning Disability	1	0.99	1	1.64	.	0.00
Multiple Disabilities	8	7.92	5	8.20	3	6.12

4 The completion rates for science and social studies for the elementary and middle school forms were lower due to sampling of participation in these content areas for two grade-level groups for each form (i.e., students were administered either science or social at these grade levels).

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
Other Health Impairment	6	5.94	5	8.20	3	6.12
Orthopedic Impairment	3	2.97	3	4.92	3	6.12
Speech Language Impairment	80	79.21	35	57.38	16	32.65
Traumatic Brain Injury	.	0.00	.	0.00	1	2.04
Visual Impairment	9	8.91	2	3.28	1	2.04
TOTAL	101	100.00	61	100.00	49	100.00

Comparing the attained video rater (VR) sample to the identified sample (see Exhibit 37), the following statements can be made:

By Form

- The attained sample approximates the expected number of students for each form: Elementary students make up 47.9% of the sample, middle school students make up 28.9% of the sample, and high school students make up 23.2% of the sample.

By IEP Disability Code

The first four rows show the primary disabilities of severe, moderate and mild mental disability and autism. If any of the mental disabilities was coded together with autism, then only the mental disability is reported. Subsequent rows show additional disabilities coded by the test administrators. Since multiple disability codes per student are permitted, the percentages do not add up to 100.

- Severe Mental Disability was sampled similarly to the expectation across forms (elementary school: 13.9%, middle school: 9.8%, and high school: 12.2%).
- Moderate Mental Disability was sampled at a similar rate (25.7%, 36.1%, and 36.7%) to the expectation.
- Mild Mental Disability was sampled at a similar rate (30.7%, 27.9%, and 30.6%) to the expectation.
- Autism was sampled at a higher rate than expected in elementary and middle school and a lower rate in high school (20.8%, 13.1%, and 6.1%).
- The total percentages of students in primary disability categories other than Severe, Moderate, and Mild Disability and Autism were represented at lower rates in elementary and middle school and a similar rate in high school compared to the identified sample (8.9%, 13.1%, and 14.4%).

Comparing the attained VR sample with the assessed population (see Exhibit 5), the following statements can be made:

By Other Demographic Variables

For other demographic variables, the proportions in the attained VR sample generally appear to correspond to those seen in the total assessed population when data were available.

- In the sample, African American (49%-59%),[5] African American/American Indian (0%-4%),Asian (0%–2%), Hispanic (0%–4%), White (37%–44%), White/African American (0%–2%), and Other (0%–2%) ethnicities were reported, representing the majority of ethnicities in the total assessed population. These percentages evidence

[5] The percentage range is reported across the three levels for which there are test forms—elementary, middle, and high school.

some variability around the corresponding population values as a result of the small sample size.

- Gender is distributed as approximately two to three males for each female; this ratio is greatest for the middle school form.
- "English Speaker II" (96% to 100%) in the sample reflects the percentage of students in the assessed population.
- Between 59% and 65% of students in the sample were eligible for Free Lunch, approximately the same as in the total population. A small group of students in the sample was eligible for Reduced Lunch (2% to 10%) which is similar to the assessed population (8%).
- None of the students in the attained VR sample were home schooled or migrant, and only two students were medically homebound; these results are comparable to the population, with reported rates of typically less than 1% for each of these demographic variables and never more than 2%.

The attained VR sample (Exhibit 38) appears to reasonably represent the identified sample (Exhibit 37) as well as the full population (Exhibit 39). The demographic variables of interest are present in the attained sample data within acceptable ranges of the identified sample and the assessed population.

Item Agreement Analysis

Within each grade band, the absolute difference between test administrator (TA) scores and AIR video rater (VR) scores for each item was computed. Scores that do not differ between TA and VR are noted as "equal"; scores differing by +/–1 score point are noted as "adjacent"; scores differing by more than +/–1 point are flagged as "discrepant." The agreement data are summarized by content area and grade band in Exhibit 39, where values indicate the average percentage of items falling within each agreement category for which there were valid matched responses across TAs and VRs.

Across content areas for the elementary school form, the majority of items (90% to 95%) were shown to be scored as "equal" between the TA and VR; "adjacent" ratings were the next most prevalent outcome (5% to 9%); and "discrepant" ratings were the least prevalent result for all content area areas (1% to 2%). On the middle school form, all content area areas show a pattern similar to the elementary form; "equal" categorizations account for the majority of ratings (94% to 97%); the "adjacent" category is next most prevalent (3% to 5%); and "discrepant" results account for the smallest proportion of ratings (0% to 1%). Across content areas on the high school form, "equal" ratings again account for the largest proportion (94% to 95%), "adjacent" is the next most prevalent (4% to 5%), then "discrepant" (1%).

Exhibit 39: Average Item Agreement Statistics by Grade Band and Subject

Subject	Agreement	Elementary	Middle	High
ELA	Equal	92.70	96.14	93.73
	Adjacent	6.12	3.31	5.40
	Discrepant	1.18	0.54	0.87
Math	Equal	89.72	95.06	95.15
	Adjacent	8.78	4.55	3.93

Subject	Agreement	Elementary	Middle	High
	Discrepant	1.50	0.39	0.92
Science	Equal	90.18	94.31	94.26
	Adjacent	8.64	4.41	5.00
	Discrepant	1.18	1.28	0.74
Social Studies	Equal	94.67	96.67	x
	Adjacent	4.64	2.94	x
	Discrepant	0.69	0.39	x

Classification Consistency Analysis (as distinct from scoring consistency just discussed in the previous section)

The reported performance levels for each student are derived from a scale score to performance level conversion process. Scale scores are produced based on conversions from the raw scores assigned by the TA. From these scale scores, students were assigned to one of four performance levels (i.e., Levels 1, 2, 3, or 4) within each grade band and content area assessment. Using the VR item scores, correspondence between reported (TA) performance levels and VR performance levels was assessed according to the kappa and weighted kappa coefficients. In ELA, mathematics, science, and social studies, consistency is assessed through *weighted kappa* (Agresti, 1990; Spitzer, Cohen, Fleiss, & Endicott, 1967), which is appropriate for ordered categories:

$$\kappa_w = \frac{\sum\sum w_{ij}\pi_{ij} - \sum\sum w_{ij}\pi_{i+}\pi_{+j}}{1-\sum\sum w_{ij}\pi_{i+}\pi_{+j}},$$

where i is the category assigned by the TA, j is the category assigned by the VR, $w_{ij} = 1-(i-j)^2/(I-1)^2$ are the weights, π_{ij} is the probability of being classified as ij, and "+" indicates agreement between categories. Kappa equals 0 when the agreement is that expected by chance; and kappa equals 1 when there is perfect agreement among raters.

Under the current condition, it must be noted that not all cases included in this analysis contained complete data. Exhibit 40 indicates the *effective sample size* ("n"; cases with information used in the content area-by-form calculation) as well as the *total count* (indicating students assigned to the current test form both with or without data for the specified content area).

Exhibit 40: Agreement Statistics by Subject and Grade Band

Subject	Statistic	Elementary (n_{TOTAL} = 101)	Middle (n_{TOTAL} = 61)	High (n_{TOTAL} = 49)
ELA	n	98	61	46
	*kw**	0.940	0.974	0.979
	95%CI	0.898 - 0.983	0.943 - 1.000	0.955 - 1.000
Math	n	97	61	47
	kw	0.912	0.956	0.968
	95%CI	0.859 - 0.965	0.917 - 0.995	0.937 - 0.999
Science	n	61	43	45
	kw	0.942	0.964	1.000

Subject	Statistic	Elementary (n_{TOTAL} = 101)	Middle (n_{TOTAL} = 61)	High (n_{TOTAL} = 49)
	95%CI	0.892 - 0.991	0.927 – 1.000	1.000 - 1.000
Social Studies	n	64	31	
	kw	0.927	0.949	
	95%CI	0.880 - 0.974	0.897 - 1.000	

Summary

TA and VR assignments of students to performance levels typically show high levels of agreement, as weighted kappa typically ranges from 0.91 to 1.00. Further, the 95% confidence intervals show that, while sample sizes for the current calculations may be small, the agreement indices are significantly greater than chance agreement and often approach 1.00. Based on the current evidence, we can conclude that the SC-Alt was accurately scored.

Appendix D: Descriptions of Achievement Levels (DALs)

Exhibit 41: English Language Arts Descriptions of Achievement Levels

Performance Level	ELA Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in reading, writing, and communication.	Students performing at level 1 should be able to • listen (as evidenced by facial expressions, gestures, or sounds) to a variety of text read aloud; • point or eye gaze to objects, pictures, or letters to complete a writing activity; • engage (using facial expressions, gestures, or sounds) in conversations focused on objects in the immediate surroundings; • listen (as evidenced by facial expressions gestures or sounds) to a speaker.	Students performing at level 1 should be able to • point or eye gaze to objects or pictures related to a variety of grade appropriate or adapted text focused on concrete concepts, read aloud; • point or eye gaze to objects, pictures, or letters to create a simple composition; • engage in conversations focused on events in the immediate surroundings as evidenced by facial expressions, gestures, or sounds; • listen to a speaker as evidenced by facial expressions, gestures, without interrupting.	Students performing at level 1 should be able to • listen to a variety of grade appropriate/adapted texts read aloud as evidenced by facial expressions, gestures, or sounds; • point or eye gaze to objects, pictures, or letters to complete more complex written products; • engage in conversations focused on objects or events outside the immediate surroundings as evidenced by facial expressions, gestures, or sounds; • listen and respond to a speaker.
2	Students performing at level 2 demonstrate foundational academic skills and competencies in reading, writing, and communication.	Students performing at level 2 should be able to • tell or show what a grade appropriate or adapted text, which contains high frequency words, is about; • identify individual words/picture symbols; • identify story elements (e.g., main idea, events, setting, and characters); • use oral and written language to describe; • select from a list of topics to generate ideas for written communication; • listen to a speaker without interrupting; • respond appropriately in conversations .	Students performing at level 2 should be able to • tell or show what text that requires only literal interpretation is about (using objects, pictures, or words); • read a variety of grade appropriate/adapted texts (e.g., recipes or advertisements); • identify story elements (e.g., main idea, events, setting, characters, and conflict); • make connections within and between texts; • use oral and written language to explain; • select from a list of topics to generate multiple ideas for written communication; • focus attention on a speaker and listen without interrupting; • engage in conversations by answering direct questions about familiar situations;	Students performing at level 2 should be able to • tell or show what a grade-appropriate or adapted text that requires simple inferences is about • read a variety of texts (e.g. recipes, advertisements, schedules, and newspapers) • identify story elements (e.g., main idea, events, setting, characters, conflict, and plot); • gather meaning from graphic representations; • use oral and written language to explain, inform, and describe; • generate ideas for written communication; • edit own writing; • engage in conversations by answering direct questions about the immediate environment or other familiar surroundings.

Performance Level	ELA Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
			• follow oral and/or written one-step directions.	
3	Students performing at level 3 demonstrate increasing academic skills and competencies in reading, writing, and communication.	Students performing at level 3 should be able to • identify story elements in text (e.g., characters, settings, events, cause and effect, and problem solution); • read words and simple sentences; • generate an idea and use words, pictures, or oral language to write; • follow one-step oral or signed directions; • communicate agreement or disagreement appropriately.	Students performing at level 3 should be able to • identify and recall details in text including main idea and characters; • draw conclusions and make simple predictions and inferences about the text; • determine meaning of unfamiliar words; • generate multiple ideas by selecting from a list and use words, pictures, or oral language to write; • initiate conversation about immediate surroundings.	Students performing at level 3 should be able to • make connections with text (plot, characters, setting); • make inferences about events in text; • understand multiple meanings of words; • compare and contrast story elements from different stories; • discriminate fact from fiction; • generate an idea and use words, pictures, or oral language to write; • follow directions to complete a task; • initiate conversations about immediate surroundings or other familiar topics.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in reading, writing, and communication.	Students performing at level 4 should be able to • identify story elements such as the main idea and cause and effect; • draw conclusions and make predictions about text; • read and understand the main idea of a simple paragraph; • create and edit personal written products; • follow two-step oral or signed directions; • take turns appropriately during conversation or discussion.	Students performing at level 4 should be able to • recognize and recall details in text, including the main idea, plot, characters, and setting; • draw conclusions and make predictions and inferences about the text; • read and understand the main idea of a simple paragraph; • explain word meanings; • create and edit personal written products; • follow oral/signed or written directions; • initiate and retell conversations.	Students performing at level 4 should be able to • recognize and recall details in text, including the main idea, plot, characters, and setting; • draw conclusions, and make predictions and inferences about the text; • read and understand the main idea of a short story; • use context clues to understand the meaning of unknown words; • make connections within and between texts and to prior knowledge, other texts, and the world; • create and edit personal written products; • use graphic representations as sources of information.

Exhibit 42: Mathematics Descriptions of Achievement Levels

Performance Level	Mathematics Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in mathematics.	Students performing at level 1 should be able to • manipulate one concrete object; • observe that two geometric figures have the same attributes; • recognize attributes of objects, such as length and weight.	Students performing at level 1 should be able to • recognize the concept of one in counting objects; • recognize that two geometric figures have the same attributes; • recognize attributes of objects, such as length, weight, and size/volume.	Students performing at level 1 should be able to • recognize the concept of one more in counting objects; • match geometric figures that have the same attributes; • respond to positional concepts such as on top of or under, off-on, above and below; • match objects by one attribute such as length, weight, and size/volume.
2	Students performing at level 2 demonstrate foundational academic skills and competencies in mathematics.	Students performing at level 2 should be able to • count objects in a set; • identify objects by one attribute (color, size, shape); • classify two - and three-dimensional concrete objects according to one attribute; • recognize positional concepts (on/off); • identify measurement tools, including graphs.	Students performing at level 2 should be able to • add and subtract using concrete objects; • sort objects by one attribute (color, size, shape); • recognize and demonstrate understanding of positional concepts (on/off, below/above); • use nonstandard units to measure; • match the correct tool to a specific task (i.e. measure length, weight, time); • identify parts of a chart, graph, or table.	Students performing at level 2 should be able to • solve addition and subtraction problems; • Identify operations (+ or -); • tell which has more in a set; • identify a repeating relationship (pattern); • sort and classify objects by one attribute, (length, height, weight. volume); • use a graph or chart to gain information.

Performance Level	Mathematics Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
3	Students performing at level 3 demonstrate increasing academic skills and competencies in mathematics.	Students performing at level 3 should be able to • demonstrate addition and subtraction concretely or symbolically; • count and compare objects in a set; • sort and classify objects by attribute (shape, size); • identify three-dimensional shapes (cube, sphere, cylinder); • use nonstandard units to measure; • find answers to questions in a graph.	Students performing at level 3 should be able to • identify the answer to one-digit addition and subtraction problems; • identify a set as having more, fewer, or the same number as another set; • extend a repeating pattern; • compare objects by attribute; • interpret information displayed in a table.	Students performing at level 3 should be able to • identify the process for solving an addition or a subtraction problem; • identify and use operational symbols correctly; • estimate the number of objects in a set; • add to find value of a set of coins; • describe, create, and complete a repeating pattern; • use and organize data to create charts, graphs, and tables.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in mathematics.	Students performing at level 4 should be able to • demonstrate understanding of addition and subtraction; • generate a pattern using three-dimensional shapes (cube, sphere, cylinder); • compare objects by attribute (length, size); • interpret information displayed in a graph.	Students performing at level 4 should be able to • solve addition and subtraction facts without regrouping; • describe and extend a repeating pattern; • interpret information displayed in a graph; • use data to create tables.	Students performing at level 4 should be able to • identify, compare, and construct numbers; • use operation symbols (more than less than and equal to) to solve problems; • add to find the value of a set of two or more coins; • identify, describe, create, extend, and complete a repeating pattern; • describe events as more likely or less likely to occur; • use and organize data to create and interpret graphs.

Exhibit 43: Science Descriptions of Achievement Levels

Performance Level	Science Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in science.	Students performing at level 1 should be able to use their senses to • observe the outcome of a simple science investigation; • sequence growth patterns; • observe and record daily weather conditions; • recognize the sun and moon and relate them to day and night; • recognize that objects move when force is applied..	Students performing at level 1 should be able to use their senses to • chose a question (how) (what if) to conduct a scientific investigation; • identify major body parts of animals; • identify the sun and moon; • recognize that objects move when force is applied and recognize speed (fast and slow); • sort by one attribute.	Students performing at level 1 should be able to use their senses to • choose questions to conduct a simple scientific investigation; • recognize that objects move when force is applied; • recognize that an object at rest does not move; • identify physical properties of matter (e.g., freezing/melting)
2	Students performing at level 2 demonstrate foundational academic skills and competencies in science.	Students performing at level 2 should be able to • generate a question to conduct a simple scientific investigation; • sort organisms by physical characteristics; • identify daily weather conditions; • recognize the pattern of day and night; • identify the position of objects such as above/below, inside, or on top; • sort materials by observable properties.	Students performing at level 2 should be able to • carry out a simple scientific investigation to answer a question; • sort and describe materials by observable properties; • sort and identify organisms by physical characteristics; • identify patterns of day and night; • recognize that an object at rest moves when force is applied.	Students performing at level 2 should be able to • carry out a simple scientific investigation related to electricity or force and motion to answer a question; • compare magnetic and non-magnetic objects; • identify the force that makes an object move; • recognize physical changes in matter; • recognize physical properties of matter.
3	Students performing at level 3 demonstrate increasing academic skills and competencies in science.	Students performing at level 3 should be able to • select appropriate tool for gathering data; • carry out a simple scientific investigation; • classify events in sequential order; • distinguish between living and nonliving things; • identify major organs of animals; • use a graph to compare daily changes in weather conditions.	Students performing at level 3 should be able to • . conduct and analyze the results of a simple scientific investigation • use graphs, tables, and charts to record data and report on the results of an investigation; • compare the characteristics of living and nonliving things; • identify what plants need to grow; • use a graph or chart to compare weather conditions each season. • classify organism into major groups.	Students performing at level 3 should be able to • predict the outcome of a simple investigation and compare the results with the prediction; • compare factors that affect an electromagnet; • identify electricity as a source of energy; • relate the change in force to the change in speed; • recognize the physical properties of two or more objects.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in science.	Students performing at level 4 should • plan and conduct a simple scientific investigation; • identify major organs of animals and their functions; • identify living and nonliving things in	Students performing at level 4 should be able to • plan, conduct, and carry out a simple scientific investigation; • communicate simple conclusions using tables and graphs; • identify simple machines (incline plane, lever,	Students performing at level 4 should be able to • plan, conduct, and analyze the results of a scientific investigation; • draw simple conclusions from distance/time graphs or tables;

Performance Level	Science Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
		terms of a food web; • identify natural resources as renewable or nonrenewable; • compare heat and light changes from season to season using a graph. • Draw simple conclusions from tables, graphs and charts	pulley); • compare data on temperature changes over time using a graph; • use a graph to show how heat and light change from season to season; • identify sources of light.	• demonstrate how simple machines are used to help people (inclined plane, lever, pulley, etc.); • predict the effect of the change in force on an object; • identify water as solid, steam, or liquid.

Exhibit 44: Social Studies Descriptions of Achievement Levels

Performance Level	Social Studies Achievement Level Definitions	Grades 3–5	Grades 6–8
1	Students performing at level 1 demonstrate emerging academic skills and competencies in social studies.	Students performing at level 1 should be able to • identify self from others; • respond to a person in authority in the home or school; • follow class rules; • engage in turn-taking; • listen to information about South Carolina history.	Students performing at level 1 should be able to • identify self from others; • respond to familiar authority figures; • follow class rules; • engage in turn-taking and sharing; • listen to information presented about significant and historical events in South Carolina.
2	Students performing at level 2 demonstrate foundational skills and competencies in social studies.	Students performing at level 2 should be able to • identify characteristics such as gender that help identify self in relation to others; • match workers to different jobs in the community; • recognize people in authority and follow class rules; • match the people we honor on some national holidays (e.g., George Washington, Martin Luther King, Jr.) with the holidays; • distinguish between past and present (match jobs of the past with jobs of the present); • match significant historical figures such as Thomas Edison to their accomplishments.	Students performing at level 2 should be able to • identify surroundings (e.g., classroom, school); • match different people to their jobs in the community; • identify people in authority and follow class rules; • demonstrate understanding of rules; • identify the people we honor on some national holidays (e.g., George Washington, Martin Luther King, Jr.); • identify the purpose of money; • match changes over time to the past and present such as communication.
3	Students performing at level 3 demonstrate increasing skills and competencies in social studies.	Students performing at level 3 should be able to • understand the concept of past and present; • demonstrate respect for people in authority; • identify major symbols of the United States; • identify why we celebrate the national holidays; • recognize that when we work we earn money to buy things; • identify features on a map of South Carolina (river, mountain, ocean); • answer questions about significant events related to the Civil War; • identify historical figures such as Thomas Edison, Alexander Graham Bell, etc. to their accomplishments.	Students performing at level 3 should be able to • identify members of the larger community (e.g., police officers, fire-fighters, doctors); • demonstrate understanding of consequences of not following the rules; • identify examples of good citizenship such as honesty, courage, etc.; • identify symbols of the United States (e.g., the flag, bald eagle); • demonstrate an understanding that we work to earn money and use money to buy things; • identify changes over time such as in travel, farming, etc.; • gain information from maps, charts, and graphs; • answer questions about key historical figures and significant historical events including the civil rights movement.

4	Students performing at level 4 demonstrate and apply academic skills and competencies in social studies.	Students performing at level 4 should be able to • place personal history on a time line; • identify the roles of leaders and officials in local government (e.g., principal, mayor, governor); • identify individuals who embody qualities of good citizenship; • identify examples of respect and fair treatment; • recognize that we exchange money for goods and services; • use a key to locate geographic features on a map of South Carolina; • answer questions about key concepts related to the Civil War; • answer questions about the accomplishments of key historical figures such as Thomas Edison, Alexander Graham Bell, etc.	Students performing at level 4 should be able to • place personal and family history on a time line; • identify roles of leaders and officials in local government (e.g., principal, mayor, governor); • identify examples of the qualities of courage and patriotism; • identify examples of respect and fair treatment and their opposites; • recognize how the amount of money available determines what we can buy; • gain information from maps and charts; • identify the accomplishments of Civil Rights leaders including Rosa Parks.

Appendix E: Summary of Linking Design

How South Carolina Alternate Assessment Standards and Measurement Guidelines (ASMGs) Overlap across Grade Bands

Common threads across grade bands were targeted in the development of some ASMGs to promote consistent instruction across the curriculum from grade band to grade band. The difference in the essence of some standards is subtle in a number of academic standards across grade bands, and in fact some standards are exactly the same in some of the content standards.

Structure of the tasks

- Each task has four to eight items. The student responses to each item are scored from one to four points depending on demands of the response.
- Only one beginning item in the first three to five tasks in an academic content area is an engagement item. Each of the engagement items is aligned with the academic content standards through the ASMGs. The remaining items of the tasks are aligned to the academic content standards through the ASMGs at complexity levels ranging from low to high. Since every student must respond to all items in at least a minimum of five tasks, every student must respond to items that assess his or her knowledge of content and skills at the grade band to which he or she is assigned.
- Items and tasks progress upward in complexity and difficulty across the performance levels at the assigned grade band.

Structure of linking tasks

- ASMGs from adjacent grade bands were examined for common threads linked to content across the two grade bands for use in developing linking tasks.
- For the mathematics assessment, several low-complexity tasks linked across all three grade bands.
- Some items were developed specifically to link to ASMGs that were common in academic demand across grade bands. Other items were designed specifically to assess only the ASMG content for a specific grade band.

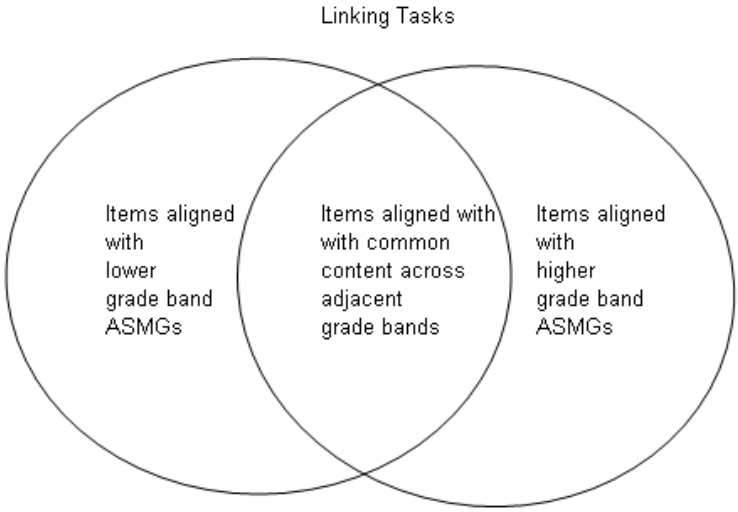

Linking Tasks
Items aligned with lower grade band ASMGs
Items aligned with with common content across adjacent grade bands
Items aligned with higher grade band ASMGs

Exhibit 45: Summary of Linking Design Across Subjects and Grade-Bands

Subject	Grade Band	Number of Items	Number of Tasks	Starting Positions									
				Starting Task 1		Within Grade-Band Linking		Starting Task 3		Within Grade-Band Linking		Starting Task 6	
				Items	Tasks	Items	Tasks	Items	Tasks	Items	Tasks	Items	Tasks
ELA	3–5	68	12	32	5	19	3	38	7	19	4	36	7
	(Linking)	*29*											
	6–8	65	12	31	5	17	3	39	7	22	4	34	7
	(Linking)	*24*											
	10	64	12	28	5	16	3	37	7	21	4	36	7
Math	3–5	53	12	23	5	14	3	30	7	16	4	30	7
	(Linking)	*37*											
	6–8	55	12	23	5	14	3	31	7	17	4	32	7
	(Linking)	*29*											
	10	60	12	24	5	15	3	34	7	19	4	36	7
Science	3–5	58	12	26	5	16	3	35	7	19	4	32	7
	(Linking)	*25*											
	6–8	60	12	28	5	17	3	36	7	19	4	32	7
	(Linking)	*20*											
	10	56	12	26	5	15	3	33	7	18	4	30	7
Social Studies	3–5	57	13	24	6	16	4	32	8	20	5	37	8
	(Linking)	*24*											
	6–8	55	13	24	6	16	4	33	8	21	5	35	8

Appendix F: Statistics Summaries for the 2009 Spring Embedded Field-Test Items

Exhibit 46: Elementary School English Language Arts (ELA) Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1406	69	0.64	3.49226	0.00	0.01	-A	-A
1407	70	0.72	0.93053	0.00	0.05	+A	-A
1408	71	0.55	0.73602	0.00	0.05	-A	+A
1409	72	0.60	0.84039	0.00	0.05	-A	-A
1410	73	0.59	0.81441	0.00	0.05	-A	-A
1355	74	0.11	1.02991	0.00	0.03	-A	-A
1357	75	0.58	1.45400	0.00	0.04	+A	-A
1358	76	0.64	1.31363	0.00	0.04	+A	-A
1359	77	0.63	1.34504	0.00	0.03	+A	-A
1360	78	0.51	1.51008	0.00	0.05	-A	-A
1361	79	0.62	1.25455	0.00	0.05	+A	-A
1308	80	0.14	1.25128	0.00	0.04	-A	-A
1398	81	0.50	0.87269	0.00	0.04	-A	+A
1304	82	0.43	1.10082	0.00	0.03	-A	-A
1311	83	0.40	0.80294	0.01	0.04	-A	+A
1312	84	0.49	1.43451	0.00	0.03	+A	-A
1313	85	0.35	0.81704	0.00	0.04	-A	+A
1314	86	0.31	1.09861	0.00	0.04	-A	-A

Exhibit 47: Middle School English Language Arts (ELA) Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1411	66	0.63	3.50686	0.00	0.01	+A	-A
1431	67	0.64	0.96039	0.00	0.03	-A	-A
1412	68	0.68	0.96034	0.00	0.03	-A	-A
1413	69	0.58	0.95177	0.00	0.03	-A	-A
1414	70	0.69	0.81009	0.00	0.04	+A	+A

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1415	71	0.65	0.81760	0.00	0.05	+A	+A
1416	72	0.58	0.77444	0.00	0.05	-A	+A
1286	73	0.46	1.44891	0.00	0.02	-A	-A
1392	74	0.64	1.33044	0.00	0.02	-A	+A
1393	75	0.60	1.18587	0.00	0.02	+A	+A
1394	76	0.64	1.28353	0.00	0.03	-A	+A
1391	77	0.61	1.28384	0.00	0.03	+A	+A
1396	78	0.48	1.39540	0.00	0.03	-A	+A
1344	79	0.54	1.43363	0.00	0.01	-A	+A
1345	80	0.37	1.43556	0.00	0.02	+A	-A
1346	81	0.62	1.04348	0.00	0.03	-A	+A
1347	82	0.41	1.40426	0.00	0.02	+A	-A
1349	83	0.40	0.93523	0.00	0.03	+A	-A

Exhibit 48: High School English Language Arts (ELA) Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1284	65	0.86	3.66867	0.00	0.01	+A	+A
1292	66	0.76	3.66867	0.00	0.00	-A	+A
1297	67	0.74	0.94737	0.00	0.06	-A	+A
1299	68	0.66	0.69565	0.00	0.05	-A	-A
1300	69	0.43	0.76087	0.00	0.05	-A	-A
1301	70	0.23	0.63323	0.00	0.05	-A	+A
1316	71	0.31	1.17572	0.03	0.01	-A	-A
1317	72	0.37	1.29713	0.03	0.01	-A	-A
1318	73	0.57	1.31511	0.03	0.02	+A	-A
1320	74	0.57	1.59032	0.02	0.02	-A	+A
1321	75	0.38	0.80520	0.02	0.04	-A	+A
1322	76	0.54	0.96689	0.02	0.04	-A	+A

Exhibit 49: Elementary School Mathematics Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1283	54	0.71	3.45445	0.00	0.01	-A	+A
1285	55	0.42	0.84765	0.00	0.03	+A	-A
1288	56	0.50	0.82154	0.00	0.05	+A	-A
1289	57	0.54	1.26899	0.00	0.07	-A	+A
1290	58	0.68	1.33883	0.00	0.08	-A	-A
1293	59	0.65	0.63533	0.00	0.07	-A	-A
1270	60	0.43	1.40379	0.00	0.02	-A	+A
1274	61	0.56	1.19093	0.00	0.03	-A	-A
1276	62	0.58	1.50106	0.00	0.03	+A	-A
1277	63	0.17	0.48617	0.00	0.03	-A	-A
1279	64	0.27	0.53526	0.00	0.03	-A	+A
1280	65	0.31	0.60498	0.00	0.03	-A	-A
1281	66	0.28	1.09403	0.00	0.05	-A	-A

Exhibit 50: Middle School Mathematics Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1309	56	0.37	0.76693	0.02	0.03	-A	-A
1332	57	0.52	1.43778	0.02	0.01	-A	-A
1338	58	0.58	1.46274	0.02	0.02	-A	-A
1339	59	0.54	1.16240	0.02	0.02	+A	+A
1340	60	0.30	1.07915	0.02	0.02	+A	+A
1342	61	0.53	1.39799	0.02	0.02	+A	+A
1343	62	0.58	1.26228	0.02	0.03	+A	-A
1363	63	0.48	0.75593	0.01	0.01	+A	-A
1364	64	0.52	1.36836	0.01	0.01	+A	+A
1365	65	0.28	1.23077	0.01	0.01	-A	+A
1366	66	0.43	1.11565	0.01	0.01	+A	+A
1367	67	0.39	1.32350	0.01	0.01	-A	-A
1368	68	0.55	0.95728	0.01	0.02	-A	-A

Exhibit 51: High School Mathematics Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1268	61	0.50	1.23052	0.00	0.03	+A	-A
1269	62	0.48	1.33442	0.00	0.04	+A	+A
1272	63	0.47	1.10390	0.00	0.03	+A	-A
1282	64	0.46	0.97403	0.00	0.04	+A	+A
1294	65	0.55	1.34853	0.00	0.03	+A	+A
1295	66	0.52	1.11111	0.00	0.05	+A	+A
1296	67	0.46	1.07541	0.00	0.05	+A	+A
1302	68	0.65	1.54305	0.00	0.03	-A	-A
1303	69	0.65	1.42384	0.00	0.01	+A	+A
1305	70	0.57	1.49000	0.00	0.03	-A	-A
1306	71	0.62	1.43624	0.00	0.04	+A	+A
1307	72	0.22	1.14189	0.00	0.02	-A	-A
1310	73	0.38	0.95222	0.00	0.03	+A	+A

Exhibit 52: Elementary School Science Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1384	59	0.68	3.52041	0.00	0.01	-A	-A
1385	60	0.56	0.71554	0.00	0.05	+A	-A
1386	61	0.68	3.51395	0.00	0.02	-A	-A
1387	62	0.58	0.66322	0.00	0.05	+A	+A
1390	63	0.45	0.52148	0.00	0.06	-A	+A
1261	64	0.10	0.58133	0.00	0.02	+A	+A
1266	65	0.46	0.62332	0.00	0.05	-A	+A
1262	66	0.54	0.57229	0.00	0.02	-A	+A
1263	67	0.57	0.63238	0.00	0.04	+A	-A
1264	68	0.35	0.66667	0.00	0.03	-A	-A
1267	69	0.42	0.66616	0.00	0.03	+A	-A
1265	70	0.25	0.45231	0.00	0.04	+A	-A
1399	71	0.29	0.61789	0.01	0.02	-A	-A
1400	72	0.48	0.56661	0.00	0.03	-A	+A
1401	73	0.49	0.66129	0.00	0.03	-A	+A

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1402	74	0.35	0.62843	0.00	0.03	-A	+A
1403	75	0.47	0.72112	0.01	0.04	+A	+A
1404	76	0.58	0.66889	0.01	0.03	-A	-A
1405	77	0.30	0.72712	0.01	0.04	-A	-A

Exhibit 53: Middle School Science Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1424	61	0.69	3.52181	0.00	0.02	+A	+A
1427	62	0.51	0.54688	0.00	0.03	+A	+A
1426	63	0.48	0.68594	0.00	0.04	-A	+A
1425	64	0.69	3.46719	0.00	0.02	+A	-A
1429	65	0.63	0.65978	0.00	0.06	+A	+A
1428	66	0.61	0.76153	0.00	0.04	+A	+A
1430	67	0.30	0.61085	0.00	0.05	+A	-A
1417	68	0.53	0.72007	0.00	0.01	+A	+A
1423	69	0.55	0.76375	0.00	0.01	+A	-A
1418	70	0.61	0.77796	0.00	0.02	-A	+A
1420	71	0.53	0.72727	0.00	0.02	-B	-A
1421	72	0.56	0.65529	0.00	0.01	+A	+A
1422	73	0.39	0.73719	0.00	0.02	+A	+A

Exhibit 54: Elementary School Social Studies Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1232	54	0.43	1.34049	0.01	0.03	+A	+A
1233	55	0.64	1.59231	0.01	0.02	+A	-A
1234	56	0.65	0.76117	0.01	0.03	+A	-A
1236	57	0.43	0.76389	0.01	0.02	-A	+A
1238	58	0.58	0.63215	0.01	0.03	-A	+A
1241	59	0.51	1.46440	0.01	0.03	-A	+A

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1245	60	0.58	1.30327	0.01	0.03	+B	+A
1246	61	0.70	1.22907	0.00	0.02	+A	-A
1247	62	0.61	1.65823	0.00	0.02	-A	-A
1248	63	0.65	1.57075	0.00	0.02	-A	-A
1249	64	0.45	1.02396	0.00	0.03	+A	-A
1250	65	0.45	1.57760	0.00	0.03	-A	-A
1251	66	0.34	1.25806	0.00	0.03	-A	-A
1252	67	0.34	1.33061	0.00	0.03	+A	+A

Exhibit 55: Middle School Social Studies Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1235	52	0.52	0.70376	0.00	0.03	-B	-A
1242	53	0.66	0.73159	0.00	0.02	+A	+A
1237	54	0.37	1.39280	0.00	0.02	-A	-B
1239	55	0.39	1.28197	0.00	0.03	+A	+B
1243	56	0.57	1.52623	0.00	0.02	+A	-A
1244	57	0.43	0.84959	0.00	0.03	-A	+A
1254	58	0.44	1.10201	0.00	0.03	-A	+A
1255	59	0.46	0.81575	0.00	0.02	-A	-A
1256	60	0.53	1.44482	0.00	0.02	+A	-A
1257	61	0.57	1.39196	0.00	0.02	-A	+A
1258	62	0.59	1.41468	0.01	0.03	-A	+A
1260	63	0.40	1.57461	0.01	0.02	-A	-B

Exhibit 56: English Language Arts (ELA): Field-Test WINSTEPS Item Statistics

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
145	-0.4658	1023	2588	0.04441	0.95	-0.78	1.33	2.74	ITS_ITEM_1406
146	-2.6564	986	951	0.17516	0.97	-0.15	0.67	-1.50	ITS_ITEM_1407
147	-0.375	982	750	0.0799	0.96	-1.11	0.86	-2.12	ITS_ITEM_1408

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
148	-1.1507	982	853	0.09831	0.95	-0.72	0.88	-1.08	ITS_ITEM_1409
149	-0.9199	981	825	0.09143	0.92	-1.51	0.72	-3.06	ITS_ITEM_1410
150	0.87568	976	1033	0.04703	1.72	9.90	1.86	9.90	ITS_ITEM_1355
151	0.02639	967	1454	0.04807	0.99	-0.12	0.96	-0.46	ITS_ITEM_1357
152	0.3153	965	1311	0.04596	0.91	-2.52	0.84	-2.39	ITS_ITEM_1358
153	0.14849	973	1341	0.05017	0.88	-3.16	0.80	-4.01	ITS_ITEM_1359
154	-0.1438	955	1498	0.05105	1.01	0.33	1.11	1.16	ITS_ITEM_1360
155	0.44515	956	1242	0.04452	0.89	-3.08	0.82	-2.70	ITS_ITEM_1361
156	0.44729	939	1220	0.0465	1.69	9.90	2.05	9.90	ITS_ITEM_1308
157	1.21593	932	850	0.04662	1.15	3.80	1.17	3.36	ITS_ITEM_1398
158	0.72008	946	1070	0.05113	1.16	4.01	1.16	3.66	ITS_ITEM_1304
159	1.39536	924	766	0.04751	1.30	6.92	1.37	6.71	ITS_ITEM_1311
160	0.1158	949	1380	0.04848	1.10	2.32	1.03	0.48	ITS_ITEM_1312
161	1.39315	940	777	0.04764	1.35	8.17	1.48	8.74	ITS_ITEM_1313
162	0.80525	933	1025	0.05135	1.24	5.78	1.24	5.36	ITS_ITEM_1314
163	-0.3433	936	2385	0.04875	1.03	0.45	1.98	5.76	ITS_ITEM_1411
164	-3.0106	917	897	0.23292	0.96	-0.11	0.82	-0.49	ITS_ITEM_1431
165	-3.0863	914	896	0.24308	0.98	-0.04	0.86	-0.34	ITS_ITEM_1412
166	-2.5437	918	888	0.19039	1.01	0.11	0.91	-0.23	ITS_ITEM_1413
167	-0.6007	907	755	0.09489	0.85	-2.73	0.63	-3.61	ITS_ITEM_1414
168	-0.714	899	762	0.09848	0.89	-1.80	0.68	-2.82	ITS_ITEM_1415
169	-0.3835	896	721	0.09038	0.93	-1.45	0.82	-1.83	ITS_ITEM_1416
170	0.25388	902	1333	0.05248	1.22	4.75	1.50	4.87	ITS_ITEM_1286
171	0.56406	904	1224	0.04906	0.94	-1.41	0.89	-1.30	ITS_ITEM_1392
172	0.86139	899	1091	0.04881	1.06	1.39	0.96	-0.58	ITS_ITEM_1393
173	0.57239	897	1177	0.0532	0.93	-1.83	0.83	-2.98	ITS_ITEM_1394
174	0.57421	892	1176	0.05299	0.95	-1.20	0.92	-1.37	ITS_ITEM_1391
175	0.36888	891	1274	0.05218	1.20	4.39	1.19	2.23	ITS_ITEM_1396
176	0.14992	891	1296	0.0578	1.00	0.13	1.00	0.01	ITS_ITEM_1344
177	0.32559	885	1292	0.05236	1.37	7.66	1.88	7.86	ITS_ITEM_1345
178	1.20168	878	936	0.05009	0.99	-0.35	1.00	0.05	ITS_ITEM_1346
179	0.42217	885	1254	0.05145	1.31	6.63	1.37	3.95	ITS_ITEM_1347
180	1.49766	880	823	0.0513	1.36	7.83	1.46	7.60	ITS_ITEM_1349
181	-0.6016	329	889	0.10169	0.72	-2.04	0.66	-1.15	ITS_ITEM_1284
182	-0.4254	331	887	0.09588	0.96	-0.26	2.03	2.63	ITS_ITEM_1292

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
183	-2.8225	312	306	0.41704	1.01	0.16	1.37	0.78	ITS_ITEM_1297
184	0.43945	314	224	0.13613	0.89	-1.83	0.76	-2.23	ITS_ITEM_1299
185	0.04098	317	245	0.14562	1.10	1.31	1.00	0.07	ITS_ITEM_1300
186	0.81841	314	202	0.12975	1.26	4.64	1.35	3.50	ITS_ITEM_1301
187	1.05107	310	368	0.09614	1.31	3.99	1.35	4.27	ITS_ITEM_1316
188	0.86765	309	406	0.0865	1.39	4.97	1.49	3.87	ITS_ITEM_1317
189	0.75138	307	409	0.09306	1.00	0.01	0.91	-1.04	ITS_ITEM_1318
190	0.27681	308	493	0.09323	1.05	0.62	0.99	0.01	ITS_ITEM_1320
191	1.98635	302	248	0.08736	1.37	4.68	1.50	4.56	ITS_ITEM_1321
192	1.6393	302	292	0.08442	1.11	1.49	1.10	1.05	ITS_ITEM_1322

Exhibit 57: Mathematics: Field-Test WINSTEPS Item Statistics

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
103	-0.7831	975	2441	0.04709	0.90	-1.64	0.88	-1.11	ITS_ITEM_1283
104	-1.4559	956	830	0.10154	1.05	0.76	1.10	0.80	ITS_ITEM_1285
105	-1.2543	942	802	0.09698	1.00	0.08	0.91	-0.76	ITS_ITEM_1288
106	0.13105	923	1236	0.04805	1.13	3.07	1.29	3.69	ITS_ITEM_1289
107	0.00507	911	1300	0.04815	0.86	-3.35	0.83	-1.93	ITS_ITEM_1290
108	-0.0547	920	616	0.07652	0.83	-5.88	0.74	-5.12	ITS_ITEM_1293
109	-0.1145	931	1335	0.04999	1.26	5.54	1.51	5.59	ITS_ITEM_1270
110	0.35129	925	1130	0.04751	1.08	2.06	1.06	0.97	ITS_ITEM_1274
111	-0.4959	922	1419	0.05676	0.95	-1.04	0.98	-0.20	ITS_ITEM_1276
112	0.83737	919	458	0.07301	1.31	9.90	1.41	9.64	ITS_ITEM_1277
113	0.61914	926	501	0.07279	1.20	7.22	1.26	6.17	ITS_ITEM_1279
114	0.29605	920	559	0.07405	1.11	4.03	1.15	3.17	ITS_ITEM_1280
115	0.63086	905	989	0.05155	1.36	8.42	1.41	8.26	ITS_ITEM_1281
116	-0.5305	871	691	0.09036	1.12	2.44	1.27	2.38	ITS_ITEM_1309
117	0.04847	888	1294	0.0528	1.14	3.00	1.28	2.71	ITS_ITEM_1332
118	-0.1544	879	1315	0.05769	1.02	0.37	1.11	1.37	ITS_ITEM_1338
119	0.67151	882	1045	0.04961	1.21	4.89	1.22	2.95	ITS_ITEM_1339
120	0.84833	884	968	0.05456	1.51	9.90	1.67	9.90	ITS_ITEM_1340
121	0.13995	884	1254	0.05213	1.11	2.43	1.09	1.03	ITS_ITEM_1342

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
122	0.44967	876	1131	0.04948	1.09	2.18	1.00	0.08	ITS_ITEM_1343
123	-0.3561	873	669	0.08684	1.00	0.00	1.11	1.14	ITS_ITEM_1363
124	0.07977	874	1211	0.05752	1.07	1.55	1.13	2.07	ITS_ITEM_1364
125	0.51408	872	1088	0.05242	1.56	9.90	1.79	9.90	ITS_ITEM_1365
126	0.76625	874	984	0.05651	1.25	5.48	1.24	4.70	ITS_ITEM_1366
127	0.3717	877	1166	0.05005	1.41	8.85	1.56	5.53	ITS_ITEM_1367
128	1.17807	866	829	0.0509	1.15	3.57	1.22	3.45	ITS_ITEM_1368
129	0.33223	299	379	0.09876	1.07	1.03	1.12	1.54	ITS_ITEM_1268
130	0.29352	297	411	0.08612	1.25	3.19	1.30	2.19	ITS_ITEM_1269
131	0.74242	300	340	0.08996	1.13	1.91	1.13	1.66	ITS_ITEM_1272
132	1.04075	296	300	0.08443	1.21	3.01	1.31	3.17	ITS_ITEM_1282
133	0.30219	298	414	0.08439	1.11	1.45	1.29	1.89	ITS_ITEM_1294
134	0.64773	294	340	0.10127	1.02	0.28	1.01	0.12	ITS_ITEM_1295
135	0.84827	293	328	0.08199	1.23	3.28	1.39	3.53	ITS_ITEM_1296
136	-0.3446	292	466	0.10452	0.90	-1.12	0.76	-1.81	ITS_ITEM_1302
137	0.22573	298	430	0.08399	0.93	-0.84	0.73	-1.62	ITS_ITEM_1303
138	0.02339	294	447	0.0911	1.09	1.00	0.94	-0.29	ITS_ITEM_1305
139	-0.0502	291	428	0.09896	0.90	-1.31	0.81	-1.88	ITS_ITEM_1306
140	0.72402	295	338	0.09096	1.46	5.97	1.51	5.65	ITS_ITEM_1307
141	1.19465	293	279	0.08737	1.21	2.97	1.39	4.39	ITS_ITEM_1310

Exhibit 58: Science: Field-Test WINSTEPS Item Statistics

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
130	-0.9574	680	1735	0.06029	0.93	-0.93	1.01	0.14	ITS_ITEM_1384
131	-0.435	656	488	0.09576	0.96	-0.94	0.87	-1.91	ITS_ITEM_1385
132	-0.7415	671	1722	0.05736	0.95	-0.66	1.29	1.97	ITS_ITEM_1386
133	-0.1377	652	450	0.09122	0.90	-2.57	0.80	-3.56	ITS_ITEM_1387
134	0.61953	648	352	0.08544	1.02	0.72	1.02	0.54	ITS_ITEM_1390
135	0.37746	649	386	0.08629	1.33	9.89	1.42	8.73	ITS_ITEM_1261
136	0.14674	651	417	0.08796	1.00	-0.10	0.97	-0.64	ITS_ITEM_1266
137	0.42591	650	380	0.08584	0.92	-2.76	0.89	-2.83	ITS_ITEM_1262
138	0.10088	638	418	0.08931	0.91	-2.75	0.85	-3.10	ITS_ITEM_1263

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
139	-0.049	643	438	0.09065	1.09	2.35	1.06	1.04	ITS_ITEM_1264
140	-0.0421	642	437	0.09062	1.03	0.74	0.97	-0.55	ITS_ITEM_1267
141	1.02583	637	294	0.08616	1.17	5.62	1.23	5.69	ITS_ITEM_1265
142	0.25016	602	380	0.09091	1.15	4.38	1.17	3.33	ITS_ITEM_1399
143	0.49077	608	353	0.08867	0.98	-0.67	0.96	-0.84	ITS_ITEM_1400
144	0.02269	605	410	0.09285	0.97	-0.83	0.94	-1.06	ITS_ITEM_1401
145	0.20032	607	389	0.09081	1.10	2.75	1.08	1.62	ITS_ITEM_1402
146	-0.3112	590	437	0.09991	0.98	-0.31	0.95	-0.66	ITS_ITEM_1403
147	0.05118	596	402	0.09347	0.87	-3.51	0.80	-3.77	ITS_ITEM_1404
148	-0.2242	590	429	0.09819	1.07	1.64	1.05	0.77	ITS_ITEM_1405
149	-0.733	633	1634	0.06454	0.92	-1.01	1.40	2.41	ITS_ITEM_1424
150	0.81542	627	352	0.08975	1.03	0.77	1.03	0.63	ITS_ITEM_1427
151	0.01882	616	441	0.0978	1.08	1.80	1.04	0.58	ITS_ITEM_1426
152	-0.4821	629	1596	0.06039	0.98	-0.28	1.35	2.21	ITS_ITEM_1425
153	0.21613	603	415	0.09625	0.91	-2.32	0.81	-2.82	ITS_ITEM_1429
154	-0.3638	619	479	0.10443	0.91	-1.67	0.81	-2.02	ITS_ITEM_1428
155	0.50092	609	383	0.0926	1.20	5.12	1.28	4.44	ITS_ITEM_1430
156	-0.0482	614	446	0.09904	0.99	-0.20	0.85	-1.89	ITS_ITEM_1417
157	-0.3238	613	473	0.10431	0.97	-0.63	0.81	-2.06	ITS_ITEM_1423
158	-0.4618	607	482	0.10827	0.94	-0.99	0.79	-2.07	ITS_ITEM_1418
159	-0.1065	608	450	0.10067	0.98	-0.33	0.90	-1.20	ITS_ITEM_1420
160	0.32359	612	405	0.09425	0.97	-0.87	0.89	-1.70	ITS_ITEM_1421

Exhibit 59: Social Studies: Field-Test WINSTEPS Item Statistics

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
85	0.0869	637	879	0.05959	1.23	4.38	1.17	2.08	ITS_ITEM_1232
86	-0.7507	640	1041	0.07438	0.86	-2.27	0.75	-3.00	ITS_ITEM_1233
87	-0.6011	633	497	0.10264	0.88	-2.25	0.73	-3.03	ITS_ITEM_1234
88	-0.5575	640	496	0.10088	1.01	0.26	0.92	-0.86	ITS_ITEM_1236
89	0.18729	635	410	0.09001	0.91	-2.81	0.85	-2.71	ITS_ITEM_1238
90	-0.1712	637	952	0.06252	1.05	0.94	0.94	-0.58	ITS_ITEM_1241
91	0.08167	632	840	0.06516	0.93	-1.40	0.90	-1.80	ITS_ITEM_1245

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
92	0.42019	624	780	0.05594	0.83	-3.76	0.73	-3.59	ITS_ITEM_1246
93	-0.571	622	1054	0.07077	0.90	-1.39	0.67	-2.35	ITS_ITEM_1247
94	-0.4708	624	994	0.06931	0.83	-2.97	0.67	-3.67	ITS_ITEM_1248
95	0.82085	617	646	0.05803	1.20	4.10	1.30	4.74	ITS_ITEM_1249
96	-0.4048	616	988	0.06792	1.08	1.24	1.06	0.52	ITS_ITEM_1250
97	0.2873	615	782	0.06312	1.23	4.50	1.29	4.50	ITS_ITEM_1251
98	0.2285	614	819	0.05947	1.26	4.99	1.47	5.41	ITS_ITEM_1252
99	0.08565	596	432	0.09959	0.99	-0.16	0.92	-0.97	ITS_ITEM_1235
100	-0.0779	598	448	0.10226	0.89	-2.44	0.78	-2.43	ITS_ITEM_1242
101	0.30096	602	854	0.06364	1.39	6.45	1.51	4.77	ITS_ITEM_1237
102	0.38107	593	786	0.07013	1.29	5.14	1.39	5.69	ITS_ITEM_1239
103	0.0101	600	935	0.06635	1.08	1.29	0.90	-0.80	ITS_ITEM_1243
104	1.54562	595	515	0.06183	1.36	6.56	1.61	7.77	ITS_ITEM_1244
105	0.97998	584	662	0.06053	1.37	6.72	1.50	6.12	ITS_ITEM_1254
106	-0.6417	588	489	0.11695	1.01	0.18	0.90	-0.69	ITS_ITEM_1255
107	0.22195	588	868	0.06485	1.14	2.39	1.02	0.21	ITS_ITEM_1256
108	0.36611	589	833	0.06339	1.09	1.67	1.00	0.08	ITS_ITEM_1257
109	0.16448	577	833	0.07022	0.95	-0.99	0.92	-0.98	ITS_ITEM_1258
110	-0.0212	585	922	0.06798	1.21	3.22	1.41	2.80	ITS_ITEM_1260

Appendix G: Marginal Reliability by Grade Band, Subject, and Starting Task

Exhibit 60: Marginal Reliability by Starting Task and Grade Band for ELA

Initial Task	Statistic	Elementary	Middle	High
1	N	375	269	113
	Reliability	0.9	0.872	0.896
	$\bar{\sigma}_{e^*}$	18.325	22.118	24.857
3	N	323	201	71
	Reliability	0.859	0.892	0.885
	$\bar{\sigma}_{e^*}$	12.614	10.713	10.238
6	N	537	607	222
	Reliability	0.767	0.743	0.814
	$\bar{\sigma}_{e^*}$	23.891	29.938	22.128

Exhibit 61: Marginal Reliability by Starting Task and Grade Band for Mathematics

Initial Task	Statistic	Elementary	Middle	High
1	N	370	269	116
	Reliability	0.846	0.826	0.873
	$\bar{\sigma}_{e^*}$	21.764	23.993	23.95
3	N	364	223	82
	Reliability	0.844	0.857	0.861
	$\bar{\sigma}_{e^*}$	15.896	13.493	14.583
6	N	501	586	208
	Reliability	0.761	0.767	0.833
	$\bar{\sigma}_{e^*}$	23.984	29.313	18.186

Exhibit 62: Marginal Reliability by Starting Task and Grade Band for Science

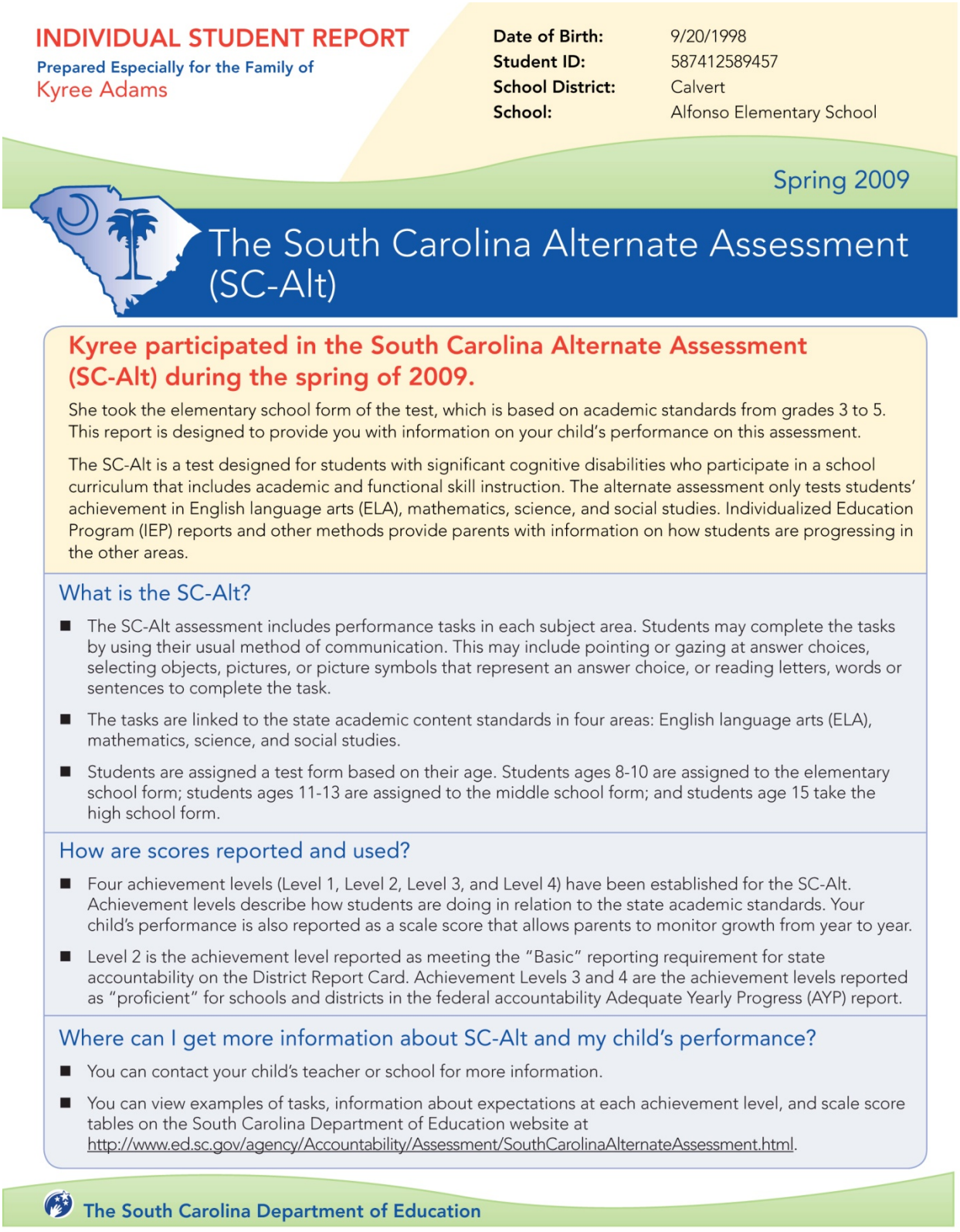
Initial Task	Statistic	Elementary	Middle	High
1	N	305	208	128
	Reliability	0.862	0.862	0.877
	$\bar{\sigma}_{e^*}$	21.614	23.378	27.771
3	N	275	145	91

Initial Task	Statistic	Elementary	Middle	High
	Reliability	0.873	0.875	0.802
	$\bar{\sigma}_{e^*}$	15.786	13.472	14.626
6	N	256	384	185
	Reliability	0.772	0.735	0.769
	$\bar{\sigma}_{e^*}$	23.195	26.781	21.553

Exhibit 63: Marginal Reliability by Starting Task and Grade Band for Social Studies

Initial Task	Statistic	Elementary	Middle	High
1	N	244	139	x
	Reliability	0.860	0.863	x
	$\bar{\sigma}_{e^*}$	23.777	18.693	x
3	N	193	114	x
	Reliability	0.844	0.895	x
	$\bar{\sigma}_{e^*}$	14.820	15.497	x
6	N	404	452	x
	Reliability	0.786	0.718	x
	$\bar{\sigma}_{e^*}$	24.283	33.347	x

Appendix H: Score Report Sample



INDIVIDUAL STUDENT REPORT

Prepared Especially for the Family of
Kyree Adams

Date of Birth:	9/20/1998
Student ID:	587412589457
School District:	Calvert
School:	Alfonso Elementary School

Spring 2009

The South Carolina Alternate Assessment (SC-Alt)

Kyree participated in the South Carolina Alternate Assessment (SC-Alt) during the spring of 2009.

She took the elementary school form of the test, which is based on academic standards from grades 3 to 5. This report is designed to provide you with information on your child's performance on this assessment.

The SC-Alt is a test designed for students with significant cognitive disabilities who participate in a school curriculum that includes academic and functional skill instruction. The alternate assessment only tests students' achievement in English language arts (ELA), mathematics, science, and social studies. Individualized Education Program (IEP) reports and other methods provide parents with information on how students are progressing in the other areas.

What is the SC-Alt?

- The SC-Alt assessment includes performance tasks in each subject area. Students may complete the tasks by using their usual method of communication. This may include pointing or gazing at answer choices, selecting objects, pictures, or picture symbols that represent an answer choice, or reading letters, words or sentences to complete the task.
- The tasks are linked to the state academic content standards in four areas: English language arts (ELA), mathematics, science, and social studies.
- Students are assigned a test form based on their age. Students ages 8-10 are assigned to the elementary school form; students ages 11-13 are assigned to the middle school form; and students age 15 take the high school form.

How are scores reported and used?

- Four achievement levels (Level 1, Level 2, Level 3, and Level 4) have been established for the SC-Alt. Achievement levels describe how students are doing in relation to the state academic standards. Your child's performance is also reported as a scale score that allows parents to monitor growth from year to year.
- Level 2 is the achievement level reported as meeting the "Basic" reporting requirement for state accountability on the District Report Card. Achievement Levels 3 and 4 are the achievement levels reported as "proficient" for schools and districts in the federal accountability Adequate Yearly Progress (AYP) report.

Where can I get more information about SC-Alt and my child's performance?

- You can contact your child's teacher or school for more information.
- You can view examples of tasks, information about expectations at each achievement level, and scale score tables on the South Carolina Department of Education website at http://www.ed.sc.gov/agency/Accountability/Assessment/SouthCarolinaAlternateAssessment.html.

The South Carolina Department of Education

Spring 2009
Kyree Adams

The SC-Alt

Mathematics

Kyree scored at Level 3 with a scale score of 480 in mathematics.

Students who score at Level 3 should be able to:

- add and subtract simple numbers;
- count and compare objects in a group;
- compare objects by color, size, or shape;
- identify three-dimensional shapes;
- read information in a graph.

	4	Students performing at Level **4** demonstrate and apply academic skills and competencies in mathematics.
Your Child's Level ►	3	Students performing at Level **3** demonstrate increasing academic skills and competencies in mathematics.
	2	Students performing at Level **2** demonstrate foundational academic skills and competencies in mathematics.
	1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in mathematics.

How you can support Kyree's learning

- Help your child add and subtract during everyday activities. For example, show her five pennies or other objects and give her one more. Then, ask her to tell you how many there are altogether.
- Play games with your child. Use games that require matching numbers (dots), such as dominoes.
- Put three objects in one pile and two objects in another pile. Ask your child which pile has more objects.
- Describe everyday household objects by shapes. For example, a can is a cylinder; a box is a rectangular prism; and a ball is a sphere.

English Language Arts

Kyree scored at Level 4 with a scale score of 495 in ELA.

Students who score at Level 4 should be able to:

- identify the main idea and make predictions about what will happen next in a story;
- write a simple story;
- follow two-step directions;
- take turns appropriately during conversations.

Your Child's Level ►	4	Students performing at Level **4** demonstrate and apply academic skills and competencies in reading, writing, and communication.
	3	Students performing at Level **3** demonstrate increasing academic skills and competencies in reading, writing, and communication.
	2	Students performing at Level **2** demonstrate foundational academic skills and competencies in reading, writing, and communication.
	1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in reading, writing, and communication.

How you can support Kyree's learning

- Encourage your child to read passages from a variety of materials (books, magazines, newspapers).
- Read a story with your child and talk to her about specific characters and the order of events in the story.
- Assist your child with writing about an event or activity using her typical method of communication.
- Encourage your child to begin conversations with family members or friends by using her typical method of communication.

2

Spring 2009
Kyree Adams

Science

Kyree scored at Level 1 with a scale score of 270 in science.

Students who score at Level 1 should be able to:

- use senses to observe the outcome of a simple scientific investigation;
- sequence growth patterns;
- observe and record daily weather conditions;
- recognize the sun and moon and relate them to day and night;
- recognize that objects move when force is applied.

4	Students performing at Level **4** demonstrate and apply academic skills and competencies in science.
3	Students performing at Level **3** demonstrate increasing academic skills and competencies in science.
2	Students performing at Level **2** demonstrate foundational academic skills and competencies in science.
Your Child's Level → 1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in science.

How you can support Kyree's learning

- Help your child plant a seed and watch it grow.
- Explain and help your child observe different weather conditions.
- Point out the moon in the sky at night.
- Show your child two balls. Roll one across a table and leave one still. Point out which ball is in motion and which is not.

Social Studies

Kyree scored at Level 3 with a scale score of 500 in Social Studies.

Students who score at Level 3 should be able to:

- understand the concept of past and present;
- demonstrate respect for people of authority;
- identify major symbols of the United States (the flag, bald eagle);
- recognize that when we work we earn money to buy things;
- identify features on a map of South Carolina (river, mountain, ocean);
- match historical figures such as Abraham Lincoln, Thomas Jefferson, etc., to their accomplishments.

4	Students performing at Level **4** demonstrate and apply academic skills and competencies in social studies.
Your Child's Level → 3	Students performing at Level **3** demonstrate increasing academic skills and competencies in social studies.
2	Students performing at Level **2** demonstrate foundational academic skills and competencies in social studies.
1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in social studies.

How you can support Kyree's learning

- Play a "Now or Long Ago" game (e.g., do we ride in a stagecoach now or long ago?).
- Find and identify items in the community such as the American flag.
- Talk about national holidays (Fourth of July and what it means) and celebrate them with your child.
- Look at a map of South Carolina with your child and find what is near the mountains and near the ocean.
- Let your child earn an allowance by doing chores or helping out by following rules. (A job can be as simple as not throwing a tantrum or allowing you to brush her teeth without fussing.)

3

Spring 2009
Kyree Adams

The SC-Alt

The following areas are tested in Mathematics:

Number and Operations
- whole numbers
- fractions
- addition and subtraction
- multiplication and division

Algebra
- patterns and their relationships

Geometry
- attributes of objects such as shape, size, color
- identification of two- and three-dimensional shapes

Measurement
- money
- length, liquid, volume, and mass and weight
- time
- equivalences

Data Analysis and Probability
- data collection and representation
- data analysis
- probability

The following areas are tested in English Language Arts:

Reading
- reading
- comprehending a variety of texts (such as fiction, nonfiction, poetry, and drama)
  Note: Reading materials may include objects, pictures or photographs, picture symbols, letters, and words.

Writing
- developing written communications (notes, stories) using the student's typical method of communication

Communication
- speaking*
- listening

**Students' typical method of communication, verbal or nonverbal, may be facilitated by using objects, pictures or photographs, picture symbols, letters and words, voice output devices, or assistive technology.*

The following areas are tested in Science:

Scientific Inquiry involves studying scientific processes and skills such as:
- observing
- classifying
- predicting what will happen in a simple scientific experiment

Life Science
- basic needs of plants and animals
- their structures and habitats

Earth Science
- weather
- objects in the sky (sun and moon)
- earth materials (rocks and soil)

Physical Science
- characteristics of objects
- the effect of force on the motion of objects
- light, heat, and electricity

The following areas are tested in Social Studies:

Social Studies Literacy Elements are concepts required for understanding this subject such as:
- distinguishing between past, present, and future
- demonstrating responsible citizenship within the school community, the local community, and national communities
- creating and using timelines
- understanding the relationship between people and the land

Academic Standards include concepts related to specific historical time frames:
- history
- geography
- political science/government
- economics

4